

Faculty of Social Sciences
University of Helsinki
Finland

CONSIDERING CHARACTER AND EXCUSES AS BLAME MITIGATION

A COMPARATIVE ANALYSIS OF STRAWSONIAN AND ARISTOTELIAN
ACCOUNTS OF MORAL RESPONSIBILITY

Eero Kaila

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Social Sciences of
the University of Helsinki, for public examination in Auditorium XIII,
University main building, on the 28 of January 2017, at 10 am.

Helsinki 2017

ISSN 2343-273X (pbk.)
ISSN 2343-2748 (PDF)
ISBN 978-951-51-2583-5 (pbk.)
ISBN 978-951-51-2584-2 (PDF)

Unigrafia
Helsinki 2017

ABSTRACT

This dissertation is aimed at clarifying the concept of *moral responsibility* within Anglophone, analytic ethics broadly defined, as well as looking at the concept of *character* to inquire about its role embedded within these theories. In this work, it is claimed that considerations of character matter when moral responsibility is assessed.

Two families of theories of moral responsibility are compared with each other starting with Peter F. Strawson's sentimentalist theory, originated in his influential article "Freedom and Resentment" (Strawson, 1962) and followed by work of others including R. Jay Wallace (1994). Scholarship on Aristotle's theory provides the basis for a similar framework of responsibility for action (*NE* III.1).

Other aspects of Aristotle's work provide additional depth to his concept of moral responsibility however. Character differs greatly in emphasis in Aristotle's theory from that of Strawson. It is claimed here that character is an essential part of human agency, and is thus a defining factor for actions taken. It is also claimed since that character is not a precise concept, taking it fully into account presents a formidable challenge to all theories attempting to explain responsibility exhaustively.

In Strawson's case, the further claim is made that what is traditionally discussed in terms of character is rephrased in terms of pleas and special conditions, which amount together into *excuses* instead. One common concept that both of these doctrines utilize is *blame*. Blame (usually accompanied with praise) is identified as a crucial component of responsibility by a majority of thinkers writing on the subject, and this reasoning is followed here as well.

An aporetic conclusion supporting critical sources is reached in terms of a common understanding of moral responsibility in Part I. The Aristotelian notion of character and the Strawsonian notion of excuse will be re-visited as examples of *blame mitigation* within the context of these theories in the systematic section of Part II. Analysis is conducted based on Bernard Williams's (1993 & 1997) explication of elements of responsibility, where a comparison is done between character and excuses appearing in the two families of theories of responsibility. Based on the findings it is concluded that the two theoretical families share similarity of structure regardless of the difference in their age, in a way that no matter whether character or excuses are used to describe the alteration of initial judgment, in terms of the end results blame mitigation appears to happen identically in all cases.

Examination of character in the context of philosophy of responsibility shows that there is room for expansion in the narrower attempts to define the concept. Comparisons of both of these theoretical alternatives are illustrated with examples and further discussion is called for.

ACKNOWLEDGEMENTS

The conception of this thesis owes a great deal to the “Responsibility: A Philosophical Enquiry into Ethics and Politics of the Concept” research project, from which gave this effort a great start, the project leader and supervisor Dr. (docent) Olli Loukola having been a constant source of assistance. I've worked most in a team with Markus Neuvonen. Our discussions as well as the conference presentations we've done together have been hugely important for this work. Liuda Kočnovaitė's philosophical insights, not to mention her organizational skills, have also been remarkable. Additionally, even though with us only for a short time, Frank Martela was present at a key moment in terms of seeking initial funding for our three-year project. He has since been a great influence as a prolific public writer as well.

Thanks to Joonas Leppänen and Sanna Tirkkonen for being good friends and co-workers as room-mates at the “oppiaine”. My senior colleagues Teemu Toppinen, Annamari Vitikainen and Simo Kyllönen, with whom I have shared work-space as well, have all done extremely well in academia, and thus have set the bar high. Additionally, Henrik Rydenfelt, Johanna Ahola-Launonen, Dina Babushkina, Malin Grahn-Wilder, Säde Hormio, Juhana Lemetti and Ville Paukkonen, to all of whom I am equally indebted, are people who have been especially efficient in organizing, presenting and giving feed-back at the various conferences, seminars and other occasions that we have met at. These thanks really extend to each and every person working on floors 5 and 6 in the A-wing of Metsätalo now and in the past. The number of helpful, brilliant people there has been too large to be possible for me to thank individually. Please accept my apologies for the inevitable omissions.

Professors Timo Airaksinen, Uskali Mäki as well as the university lecturers Olli Loukola, Heta Gylling and Arto Siitonen were in charge, while I was at the department. Professors Jaakko Kuorikoski and Aki Lehtinen have done good work recently. Long-time head of department Kristian Klockars has also been very gracious in his trust in my abilities leading to my occasional stints as an assisting teacher at the department. The invaluable commentary provided by the pre-examiners Mikael Karlsson and Jussi Suikkanen helped to clarify key features of this final version. Thanks to Ilpo Halonen and Karolina Kokko-Uusitalo for running the office without a hitch. Of professors from other institutions whom I've met, Jane L. McIntyre, Christia Mercer, Jonas Olson and Henry Rosemont Jr. have made a positive impression as being exemplary academicians.

Extra special thanks go to aforementioned colleagues, who took great pains in reading and commenting on my (awful, awful) drafts. To these people, I am eternally grateful. Thanks to Inkeri Koskinen for informing me about a work space (*suoja*). Salutations to Severi Hämäri, Tero Ijäs and Pii

Telakivi too for sharing your struggles at a writing club a while back. Grateful thanks to Markku Oksanen for advice on structure, and to Matias Slavov and Aino Lahdenranta as well for recommending me as a session chair at the Hume Society conference, where at I met countless smart and delightful people, including the author of one of the articles referred here, professor emerita McIntyre. Grateful thanks to all my friends who have given me positive encouragement throughout the years.

The research completed for this dissertation has been funded for the most part by the University of Helsinki and its various offices, including three-year funding for the above-mentioned project, personal grants from The Research Foundation of the University of Helsinki, The Fund for Humanities and Social Sciences, as well as the Faculty of Social Sciences. All of these grants have helped the work progress. Additionally, the Department of Political and Economic Studies has employed me sporadically as a teacher and as a supervisor of exams, for which I am also grateful. The Emil Aaltonen Foundation provided me with a much needed boost for the finalization period with their encouragement grant in the summer of 2016. Finally, I have enjoyed travel grants provided by the Chancellor's office in Helsinki, the former national Doctoral School of Philosophy and a very generous one from the University of Oslo.

Last, but not least, my unending gratitude goes to my wife Katja and my son Helmer. Throughout my graduate studies, Katja has been working all the while, supporting our family. My son, Helmer, naturally the most important person in my life, is doing well in school. The years shared with Katja, Helmer and Alisa the Basset Hound have been the best, and I hope for many more. My heart-felt thanks go to my family including parents Erkki and Pirjo, sisters, Ella, Marja and Anna, grandparents Helmi †, Leena and Mauno, as well as the families of aunt Raija, in-laws Paavo and Marja and sisters-in-law Kirsi and Karita.

Chapter 5 is based on scholarly work done in preparation for the previously published, peer-reviewed Finnish-language article on the topic of Aristotle's conception of responsibility in *Historiallinen aikakauskirja* 02/11 (Kaila 2011). Since its publication, the work on the article has been expanded from these efforts and presented at the World Congress of Philosophy XIII in Athens, Greece in August 2013.

Parts of the thesis have been additionally presented at the following venues: Two graduate student conferences of the former Finnish Doctoral School of Philosophy were participated in Tampere in 2011 and 2012. The topics of these were focused on moral sense theory. A general presentation of the concept of character in theories of responsibility was held at the same venue in 2013. The analysis section (chapter 7) was later presented and discussed at the University of Helsinki's Moral and Political Philosophy Research seminar in December of the same year. A paper was read at a graduate seminar in 2014 held by (the new) doctoral programme of University of Helsinki. This occasion was dedicated to the methodology section of this work, which is found in chapter 2. Finally, a presentation on Wallace's exemptions (ch. 4.1.2) was presented at the University of Turku in spring of 2016.

For Katja & Helmer

CONTENTS

Abstract.....	3
Acknowledgements.....	3
Contents.....	8
1 Introduction.....	11
2 Methodology and Scope.....	17
2.1 Methodology	19
2.2 Literature review	24
Part I: Two Theories of Responsibility	31
3 Strawsonian Responsibility	32
3.1 Responsibility in “Freedom and Resentment”	40
3.1.1 The Problem of Determinism	42
3.1.2 Reactive Sentiments and Blame	49
3.1.3 Special Considerations or Pleas.....	56
3.1.4 The Web of Human Attitudes and Feelings	59
3.2 Summary	62
4 Responses to Strawson	64
4.1 R. Jay Wallace’s <i>Responsibility and the Moral Sentiments</i>	65
4.1.1 Wallace’s Strawsonian Responsibility	66
4.1.2 Excuses and Exemptions	70
4.1.3 Reflective Self-control and the Morality System.....	74
4.2 Discussion: The Current Debate on Blame	83
4.3 Criticism.....	94
4.3.1 Herbert Fingarette: “Responsibility”	98
4.3.2 Marion Smiley: Moral Responsibility and the Boundaries of Community	100

4.3.3	Bernard Williams: <i>Shame and Necessity</i> & “Moral Responsibility and Political Freedom”	103
4.3.4	Gary Watson “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme”	110
4.4	Summary	116
5	Aristotelian Responsibility	118
5.1	Aristotle’s Theory of Responsibility	120
5.1.1	Responsibility for Action	123
5.1.2	Character, Virtues, and Moral Responsibility	133
5.1.3	The Political Responsibility of the Citizen.....	140
5.2	Discussion: Responsibility for Character in NE III.5 – Moral Responsibility or Moral Education?	144
5.3	Summary	155
	End of Part I “What does responsibility mean here?”	156
	Part II: Character and Excuses in Theories of Responsibility.....	160
6	Concepts of Character and Excuse	163
6.1	Moral Character	166
6.2	Moral Excuses: From Strawson’s Special Considerations to Watson’s Exemptions	175
6.3	Examples	190
6.4	Summary	196
7	Analysis.....	200
7.1	Aristotle’s Character-based Blame Mitigation	203
7.2	Strawson’s and Wallace’s Excuse-based Blame Mitigation	205
7.3	Findings	210
8	Concluding Statements.....	213
8.1	The Argument of this Work	213
8.2	Discussion	216

9	Conclusion	227
	Bibliography	232

1 INTRODUCTION

Difficulties in defining moral responsibility

The purpose of this work is to look at a concept embedded in the theories of moral responsibility that is seldom at the forefront of discussion. In this work, it is claimed that considerations of character matter when moral responsibility is assessed. It is also claimed since that character is not a precise concept, taking it fully into account presents a formidable challenge to all theories attempting to explain responsibility exhaustively.

Philosophy, at least in its Anglophone, analytical context, has been for the majority of the previous century occupied with the problem of determinism while trying to define responsibility. The effort put into the free will vs. determinism debate has been monumental and, with all due respect to the scholars devoting their time to the subject, it seems that the nature of professional academic philosophical conduct with different agendas and starting premises makes any satisfactory conclusion to the debate even undesirable. These troubles considered, an attempt to bring clarity to the concept of moral responsibility is made by examining contemporary academic discussion in context of moral philosophy and by comparing current understanding of the topic with historical sources.

The thesis is divided into two main parts: “I Two Theories of Responsibility” and “II Character and Excuses in Theories of Responsibility”. The first part, consisting of two different takes on moral responsibility, will frame the research area. As responsibility itself is a very broad topic, almost all-inclusive by definition, a further distinction is needed between moral responsibility, which is the topic here, and other forms of responsibility such as collective, legal and political responsibilities etc., which are not. The general principle followed is that since these related concepts cannot fully be separated from each other, even the forms of responsibility outside the narrower definitions of moral responsibility are taken into account if they provide relevant information. Later in the work, the specific concepts of moral responsibility will be further analyzed in terms of their component elements, of which the most significant is *blame*, which has received a lot of attention from philosophers in recent times.

The primary philosophical method used here is comparative conceptual analysis, supported by the history of philosophy. The study contrasts two traditions in the history of philosophy: that of philosophical sentimentalism, focusing on Strawsonian concepts of responsibility and blame and that of Aristotelianism, or more accurately the recent scholarship on Aristotle. The concept of responsibility and its nature in Anglophone analytic ethics as well as political philosophy form the background of this study. It will be analyzed in depth particularly the more specific, yet still very broad topic of moral

responsibility. The philosophy of responsibility has recently seen a shift from the more ontological questions of free will vs. determinism toward more naturalistic and pragmatic questions concerning moral psychology and society. Peter Strawson and his article "Freedom and Resentment" (2008/[1962]), the most influential contribution to this shift, serve as an anchor point for the other interpretations and developments that concern the concept.¹

Strawson's article has established the context of most attempts to define responsibility since. It distinguishes reactive sentiments (blame in terms of emotions such as resentment, indignation and guilt) as the quintessential element of inquiries into responsible action, an element which had not been debated for more than two centuries since the time of David Hume and classical sentimentalism, which was a different approach than the dominant, mainly utilitarian interpretations of responsibility. Yet the definition itself was not at all precise, with the disentangling of free will and determinism taking most of the effort. Throughout history, the topic of responsibility has attracted various considerations, including those of causal action and its response. Analysis of the concept is possible when these diverse facets are taken into account. Responsibility, conceived as the act and feedback coming back to the agent based on the act is thought of here as the default view on the subject.

Some alternate viewpoints following Strawson's theory are discussed as well as some direct criticism. Many think that this particular focus is not sufficient for it to satisfy everything a satisfactory conception of responsibility needs to encompass.² Strawson's critics, some of whom are looked at here, have very illuminating ideas on the subject. The shortcomings of this Strawsonian "default-view" of responsibility can be pointed out through these critical takes. Throughout the work of his followers such as R. Jay Wallace (esp. 1994), it is possible to question some aspects of the Strawsonian position. The main conceptions of responsibility, along with related critical discussions, are considered. How are the other viewpoints pursued here? The natural way to identify a settled discourse and its long-established principles is to ask how things were before. The history of philosophy is discussed here in order to see what has been omitted from the most common way of defining responsibility. Having something in common with each other, these criticisms look back at how responsibility was used in the past in different sets of theories. For example, Gary Watson (1987) makes valuable contributions to the subject, with the help of an empirical example in which the knowledge of the developmental history changes the way a

¹ "Freedom and Resentment" (Strawson 2008/[1962]) is referred to with the abbreviation *FR* in the text.

² See, among others, Fingarette 1966, Smiley 1992, B. Williams 1993 & 1997, G. Williams 2010.

clearly responsible criminal is viewed.³ Another one, a more pragmatic take on the concept by Marion Smiley (1992), implies that there have been two historical alternatives to the way that the concept of responsibility is discussed now. Her approach divides responsibility into two questions: that about causal responsibility and that about blameworthiness. Bernard Williams's list of the elements of responsibility is utilized here in order to compare the theories.⁴ The criticisms, along with other alternative viewpoints, together highlight a factor that has not been present in Strawsonian theories. The question is how differently has the topic been discussed in history of philosophy? These three criticisms along with their supporting statements will be looked at in the section following discussion of the Strawsonian conception of responsibility.

The study is complemented here by the consideration offered in the classic works of Aristotle.⁵ Aristotle, who is viewed here as the original theorist of responsibility, deals with the same subject matter in the *Nicomachean Ethics* (and to a lesser extent in the *Eudemian Ethics*, *Rhetoric* and *Politics*).

Core Components of Moral Responsibility

Aristotle's scholarship has produced an alternative to the concept of responsibility that is now very popular. This Aristotelian conception of responsibility relies partly on interpretation, as the concept and the philosophical discourse itself have been formulated much later.⁶ Nevertheless, there is common ground between these historically very far removed schools of thought.⁷

One common concept that both of these doctrines utilize is *blame*. Blame is identified, along with praise, as a crucial component of responsibility by a majority of thinkers writing on the subject of responsibility, and this reasoning is followed here as well. In the case of Strawson and his followers,

³ According to many Strawson commentators, his theory lacks what Gary Watson briefly illustrates as "historical responsibility". This is discussed in chapter 4.3.4. (Watson 1987, 279-280. See also McKenna & Russell 2007, 13.)

⁴ Bernard Williams's elements of responsibility, based on ideas in Williams 1993 & 1997, are discussed in chapter 4.3.3. The analysis in chapter 7 owes its framework to these elements.

⁵ The reason for this conscious oversight is also discussed: the concepts of character, historical responsibility and their variations have traditionally been part of the incompatibilist arguments for freedom of the will. Watson (1987), for example, tries to take these subjects into account separately, in order to produce novel approaches to the concept.

⁶ This is done through the works of commentators including Irwin 1980, B. Williams 1993, Broadie 1991, and Meyer 2011 et.al.

⁷ The Strawsonian conception of responsibility has been so influential, that it has also in part affected the Aristotelian conception that contemporary scholars maintain. I attempt to note these cases where there is potential for mix-ups.

responsibility is identified through blame as a response to some form of misconduct carried out by another (see McNamara, 141). Thus responsibility is strongly linked to the action of individuals and to how the *reactive sentiments* are experienced in their context.

Discussion on *praise* and *blame*, sentiments that are directed back toward the voluntary actions of an individual, goes all the way back to Aristotle. The groundwork for all subsequent discussions on responsibility is undertaken in the *Nicomachean Ethics*, primarily in book III. Aristotle's actor is equally subject to the praise and resentment that is bestowed upon him/her because of his/her actions, but in his case the community apportioning the praise or blame is more directly involved in the process. After all, Aristotle was not concerned with an instrument of ethical analysis, but rather with describing an ideal society. One element that is still connected with the discussion of responsibility and is directly followed by the talk on responsibility for action in the *Nicomachean Ethics* is the way that action reveals the character of the actor to the society. These considerations of character make determining responsibility a deeper effort than that simply checking the boxes of voluntariness, consciousness and self-control.

Aristotle's theory provides the basis for a similar framework of responsibility for action (*NE* III.1) to Strawson's. Other aspects of his work provide additional depth to his concept however. The concept of *character* differs greatly in emphasis in Aristotle's theory from that of Strawson. Character can have an effect on responsibility assessments that lead to different outcomes when an otherwise identical action is evaluated morally. This kind of alteration of judgment is present in Strawson's work too, but it is discussed under the guise of the *excuse*,⁸ which does not necessarily have to be in contact with actual actions or events in the way that Aristotle's character does.⁹

It appears that concerns about character have been left out of modern considerations of responsibility, or that it has received considerably less attention. As will be claimed later, the concept of responsibility has of

⁸ The key terms of *excuse* and *exemption* that are used here were coined by Strawson's commentator Gary Watson (1987, 259-261). Strawson uses the terms "special considerations" and "pleas" to demonstrate the cases in which concerns about the moral status of the agent can affect the response to the act. These phrases are interchangeable with the excuses and exemptions, which are introduced in part I, chapter 3.1.3. In a nutshell, the definitive concept of a *Strawsonian excuse* is found chapter four of *FR* (Strawson 2008/[1962], 7-10), despite the fact that Strawson writes about the subject using different (yet synonymous) phrases such as special considerations and pleas. The reasoning behind the use of these concepts is explained in the closer analysis in chapter 6.2 of part II.

⁹ Character, on the other hand, was still a part of the subject area of responsibility-related theories discussed in David Hume's time, but has received less attention apart from the theories of responsibility originating from the resurgence of virtue ethics by Anscombe (see 1958), MacIntyre, Foot et.al.

necessity had to be made lighter or *thinner*,¹⁰ so that a viable tool of analysis could have been utilized. Bernard Williams, one of the most famous historically minded critics of the Strawsonian conception of responsibility offers an intricate classification of the aggregate term, responsibility, into its constitutive components.

Character and Excuses as Blame Mitigation

Finally, after the work by the philosophers concerned has been discussed in the substantive section of Part I, the Aristotelian notion of character and the Strawsonian notion of excuse will be re-visited as examples of *blame mitigation* within the context of these theories in the systematic section of Part II. While the complete dissertation provides a unique viewpoint on the subject, as a distinctly new contribution is presented in the way Williams's elements of responsibility are used to compare the two theories. The comparison committed in the analysis chapter will help to allay concerns related to the difficulty in defining moral responsibility. Blame mitigation is considered here as the natural outcome of re-evaluating the initial reactive sentiment of blame that occurs in connection with the observed act requiring moral evaluation, resulting in lesser or even completely removed emotional reaction compared to the original event. These two traditions of moral philosophy appear to have something in common in considering these forms of blame mitigation. The "special considerations [which] might be expected to modify or mollify [the initial] feeling or remove it altogether", which are discussed relation to excuses and exemptions, where the former describe more closely the cases where the effect of the reactive sentiment is somewhat lessened, and exemptions where the sentiment is canceled outright (*FR*, 7), or in terms of character and its various stages of development. Keeping in mind that the classical idea of character itself is not without problems, it should not be advisable to reinstate it in some out-dated form. Rather, further discussion is called for and possible avenues of development will be considered by asking how the various means of blame mitigation within the theories of responsibility explain the concept.

I claim that *character* is an essential part of human agency, and is thus a defining factor for actions taken. It is accounted for, although insufficiently, in Strawson (see *FR*, 25-26). Strawson implies however that: (1.) character is important, but (2.) is left out of his inquiry. Nevertheless (3.) it is possible to reconstruct the concept through the moral sentiment, or, in his case, the reactive sentiments. It seems that this commitment has not been shared by the majority of Strawson's followers, who seem to share the emphasis on excusing practices. In Strawson's case, the further claim is made that what is traditionally discussed in terms of character is rephrased in terms of *excuses* instead. Further on, I argue that in Aristotle's texts and the work of his

¹⁰ To use the terms thick and thin concepts made popular by Bernard Williams (1985).

scholars, the discussion related to blame mitigation is related to his idea of character. The virtues that are supposed to be attained when the actor reaches adulthood determine whether the proper level of responsibility is achieved by the individual. Blame as an element of responsibility in Aristotle functions very similarly to that used in the later counterparts.

Based on the analysis conducted in chapter 7, I conclude that no matter whether character or excuses are used to describe the phenomenon, in terms of the end results blame mitigation appears to happen identically in all cases. Comparisons of both of these theoretical alternatives are illustrated with examples. Especially interesting comparisons are provided by the recent debates in the philosophy of responsibility, both in the Strawsonian and the Aristotelian camp. A child, for example, can be either excused on the grounds of a Strawsonian exemption, according to which being “only a child” suspends our “ordinary reactive attitudes toward the subject” (*FR*, 8-9).

Alternatively, in the Aristotelian sense we might forgo condemnation of the agent on grounds of the knowledge we have about the character of the child. It would be irrational and unfair to expect similar action from him/her. Another example that provides interesting questions about the nature of agency is the psychopath.¹¹ The interesting part is that the psychopath illustrates is that of someone who can cognitively understand moral principles and the consequences of actions related to them full well but, in the words of Watson, “they just do not care about that.” This case is interesting as it puts the theories of responsibility into the paradoxical position that on the one hand the person is exempted from responsibility, but on the other he can consciously and uninhibitedly cause harm to fellow human beings. Watson provides an interesting distinction between different “faces of responsibility” that help to understand these difficult cases conceptually (Watson 2011, 308-309; see also Watson 2004/[1996].)

The way of looking at the theories of responsibility here is related to the means by which they discuss the common element of blame. Blame mitigation is at the heart of each of the accounts examined. The thesis is that *blame mitigation* is a concern shared by different philosophical traditions, and as such is associated with the core ideas of philosophy of responsibility in general. A basic level of comparability between the main theories of responsibility is assumed, since a better understanding of the phenomenon and its implications can be gained by looking into the initial Strawsonian position as well as its criticism, and contrasting this position with Aristotelian history through commentary on it. This, by itself, is argued to contribute to the current debate on the naturalistic foundations of blame and moral responsibility as its extension.

¹¹ There has been some interesting debate on the subject by authors such as Thomas Scanlon and Gary Watson, to mention a few.

2 METHODOLOGY AND SCOPE

Throughout the history of Western philosophy, the subject of responsibility has been discussed about among many different schools of thought, with different aims and fundamental rules extending from classical Greece, Aristotle being the most important figure, to the moral sense theorists including David Hume and Adam Smith, as well as from the more modern schools of utilitarianism to Kantianism, to cover all major disciplines of philosophy, with whomever has had an interest in ethical theory. An encompassing account of the subject would effectively be a lifetime's work, and is naturally not pursued here as such. As responsibility in philosophy is a very broad topic, and it would be impossible to accommodate all viewpoints on the subject at least within the limitations of a single study, this chapter outlines what is included within this dissertation.

Basically, what is attempted here is a diverse look at the contemporary discussion on moral responsibility within Anglophone philosophy, situated in analytic ethics broadly conceived. Since the research area would still be too large defined as such, measures have been taken to keep the subject manageable. A number of important insights and points of view will certainly be lost in the process, but to minimize this risk these short-comings are alleviated somewhat with the help of work done by Garrath Williams's bibliography, "Moral Responsibility: Oxford Bibliographies Online Research Guide" (G. Williams 2010), which provides a sufficiently broad and comprehensive list of important monographs and articles. Needless to say, the present research is indebted to Williams' effort.¹² The bibliography

¹² Garrath Williams has done admirable work on cataloguing the theories of responsibility in recent years, while making his own contribution to the discussion as well. Most importantly his "Oxford Bibliography on Moral Responsibility" (2010) is fairly comprehensive and is used as a point of comparison for this literature review (ch. 2). William's stance, using Strawson's reactive sentiments approach (see ch. 3) to anchor responsibility theories, is seconded here. In addition to his introductory review, Williams has also published "Responsibility as a Virtue" (2008), "Responsibility" (2009) and "Sharing and Holding Responsible" (2013). Williams (2008) highlights an important interpretation of the concept, which is different from the usual definitions. His internet article "Responsibility" (2009) contains some valuable distinctions, including that between retrospective and prospective responsibilities. "Sharing and Holding Responsible" (2013) is an interesting article that distinguishes two themes in the discussion of moral responsibility: the usual questions of blameworthiness and retribution as well as the further question of standing to respond or reproach (see G. Williams 2013, 360, ch. 4.3). Since Williams's bibliography was published, there have been a couple of additional published works on moral responsibility that will be referred to here, including Coates & Tognazzini 2013 & 2014, as well as Shoemaker (ed.) 2013 & 2015; Shoemaker & Tognazzini (eds.) 2014; Clarke, McKenna & Smith (2015) can also be read as a continuation of this debate. Second, there are emphases in the bibliography that will differ from the preferred view-point of this work. Regarding the first

mentioned covers more areas than this study will, and it is recommended for further information and reference. The bibliography is also helpful in providing an introduction to those areas that are left out of this dissertation by necessity. There is, naturally, an enormous amount of scholarship on moral responsibility, too large to survey within the confines of this study. These omissions are regrettable, but unavoidable due to the breadth of alternative foundations from which moral responsibility has been approached throughout recent history.¹³

objection, the general philosophical debate on moral responsibility is one of the most extensively discussed. Nevertheless, there are still works being published annually on the subject, some of which fall into established disciplines (e.g., Wallace 1994, 2013), while others go for intentionally fresh attempts to tackle the subject from a novel point of view (Smiley 1992, B. Williams 1993, 1997).

¹³ Notable omissions include: Hume's theory of responsibility: While not discussed extensively, Hume's *A Treatise of Human Nature* (THN, 2000/[1739–40]) and *Enquiry Concerning Human Understanding* (2000/[1748]) are referenced here only in two important ways: first as a historical background for Strawson's work, as THN is the most renowned statement of the classical sentimentalist position. Second, Hume has a unique outlook on the role of character in relation to responsibility.

Kantianism: While Kant's influence cannot be escaped in modern philosophy, and indeed while many of the authors cited here show Kantian influences, specific works by Immanuel Kant and pure Kantianism in general are not in focus here. If Kantian responsibility had been included, in addition to his works, Christine Korsgaard's *Creating the Kingdom of Ends* (1996) would have been one of the key sources that would have been used to study Kantian moral responsibility.

Utilitarianism/"The Economy of Threats": As an early example, the utilitarian theory *economy of threats* limited the role of responsibility to the practice of avoidance. For utilitarian instrumentalism, also known as the Schlickian approach, referring to its founder Moritz Schlick (Schlick 1966/[1939]), responsibility is essentially something of a limit on liberty to act. At the time, Strawson was debating against P. H. Nowell-Smith (1948). More recent contributors have been J.J.C. Smart (1961 & 1963), D. Dennett (e.g., 2011) and R. Arneson.

Contractualism: H. L. A. Hart (1968) is noted in connection with *retribution* among other concepts for its wide-reaching influence. Similarly, the early work of Thomas Scanlon (1998), who is briefly discussed in connection with the concept of *blame* (4.2), can be read as belonging to this group.

Incompatibilism, Hard Determinism and Skepticism: The incompatibilist concept of responsibility is an alternative approach to responsibility, include C. A. Campbell (1951), R. Kane (1996), J. M. Fischer (1987, 1998 with M. Ravizza), D. Pereboom (2001), M. McKenna (1998, 2005, 2009), S. Smilansky (2000, 2001), S. Wolf (2011, 1990, 1987), A. Mele (see 2009), D. Dennett (2011) and Clarke & Capes (2013). Ginet, Chisholm, as well as P. Todd (2012). The *Stanford Encyclopedia of Philosophy* also has an article on the subject ("Incompatibilism"). An example of a moral responsibility skeptic is Bruce Waller (2011).

Harry Frankfurt's moral philosophy: Being the originator of the so-called "Frankfurt-style examples", Frankfurt's theory relies on a specific psychological mindset, according to which our desires and wants are hierarchically ordered. Frankfurt's theory is really more to do with questions about moral agency as a requirement for responsibility than about an actual concept of responsibility as such. Secondary sources referred to here that support Frankfurt's view include Doris (2002).

2.1 METHODOLOGY

Two Research Questions

The original motivation for this work stemmed from the fact that it has been seemingly impossible to find identical definitions of the concept of responsibility among the different theories of the concept. This doctoral dissertation starts from this problem, and will ultimately culminate in a comparative analysis in which the two important historical theories of responsibility, Strawsonian and Aristotelian, are compared. The chronological order, in terms of the chapters, is reversed. This is because the Strawsonian, broadly naturalistic perspective is used here as a point of reference, or a “lens” for both of the theoretical wholes.

Two main concepts, *responsibility* and *character*, will be closely examined. Both are notoriously vague and difficult to define. Both have been studied extensively from many different sources.¹⁴ Neither has received a conclusively and unanimously accepted definition, and as such they have remained continuing subjects of academic debate. While responsibility is the primary concept here for scholarly discussions, to which the thesis is connected, character is the one of what is learned more about.

During the research process, a more specific problem than the definition of responsibility presented itself. What is the role of character in the theories of moral responsibility? Character in relation to responsibility, it is argued here, while receiving some attention in philosophy, is insufficiently understood. Character is an important and even an inalienable concept *within* the theories attempting to define responsibility and moral agency in particular. The relation between character and moral responsibility is

Philosophy of Law, Collective Responsibility and CSR: Christopher Kutz’s article “Responsibility” (2002) in *Oxford Handbook of Jurisprudence and Philosophy of Law* has gained a lot of popularity, and along with Scanlon, the more relational aspects of responsibility have been brought back into the discussion. The work of Michael S. Moore (incl. 1990) is also noted for his contribution in this area. There is a very vibrant discussion going on related to collective responsibility as well as corporate social responsibility that deserves attention. See Mäkelä (2013) which emphasizes Marina Oshana’s theory of responsibility. Another gateway to the subject is the *Stanford Encyclopedia of Philosophy* article “Collective Responsibility” by Marion Smiley (2010).

¹⁴ David Shoemaker captures the elusiveness of the themes of agency and moral responsibility at the beginning of his edited series *Oxford Studies in Agency and Responsibility* by referring to the wide variety of disciplines considered in connection with questions on the topic (including, for example, the relation between responsibility and determinism and free will): “Work on these questions, while more or less having an uneasy home base in the world of moral philosophy, draws from a diverse range of cross-disciplinary sources, including moral psychology, psychology proper (experimental, developmental etc.), philosophy of psychology, communicative disorders, philosophy of law, legal theory, metaphysics, neuroscience, neuroethics, political philosophy, and more.” (Shoemaker 2013, 1-2.) The same can be said about the nature of character.

fortuitously a subject that is manageable within the bounds of a single book. This conceptual relationship is the object of this study, and its clarification is the goal. The relatively slow turn toward the second main question should be adequately reflected with a switch between the two research questions and the two parts of the thesis that they in turn reflect. The first question is a prerequisite to proper analysis of the second, and any advances and clarifications related to the second help to further define the ones related to the first as well. The different ways that the two theories of responsibility treat the concept of character compels the articulation and careful consideration of the related concepts and the relation between them.

It is thought impossible here to be able to offer comprehensive accounts of either of the two extensive historical accounts of responsibility. For this reason, an approach resembling a *heuristic* is taken toward these two accounts at the end of part I. Once the examination of the two approaches is sufficiently executed, part II will concentrate on the relevant concepts and their analysis. This book has been guided by two main research questions, which are represented by the two main parts (Part I & Part II) of the dissertation. The first research question is as follows:

Q1. How is responsibility understood by philosophers today?

The purpose of this work is to highlight an aspect of the debate on moral responsibility that is seldom at the forefront of the discussion. It is among the goals of this work to examine the most common definitions of moral responsibility in contemporary academic philosophical discussion. A comparison is made with historical sources on the study of the subject area, and an attempt made to upset the status quo based on the findings, so to speak.

In this work, the basis for the answer to Q1 is Strawson's theory of reactive sentiments, which is discussed in chapter three. Following Strawson's theory, his followers work on the same topic and some alternative viewpoints are discussed, including direct criticism. Chapter five consists of an examination of Aristotle's work under the same subject matter. These alternative accounts reveal aspects of the concept that have not been present in Strawsonian theories as much as they probably should have: on the one hand, the dimensions of historical and social implications of the concept, as well as the concept of *character*, into which the considerations of the former are often condensed. On the other hand, it is acknowledged that the idea of character itself is not without problems, and it is not wise to be reinstate it at least in its original Aristotelian form.

The works of the classics have been helpful in showing how the influence of character has been unproblematic in them, but, for various reasons, has since been left out of theories of responsibility, as Smiley, for example, points out. Q1 is ultimately answered in the form of heuristic in the end of part I. As

this conclusion leads to partial aporia, the omissions lead to the second question of this thesis:

Q2. How should character be understood in the context of the philosophy of responsibility?

Questions about the role of character and its related concepts are discussed in the second part of this work. The start of the redefinition process in character's case is to go back in history, and for this purpose the second method, that of historical scholarly study, is needed. The reason historical scholarship is relied on here is because, while conceptual analysis provides the main tool for identifying and comparing the central, contemporary theories that are relevant to the subject, a backdrop for these modern theories is needed in order for the topics to be presented intelligibly.

Part of the problem is that what philosophy discusses under the topic of character is now mostly considered to be of concern to the empirical sciences of psychology, sociology, and so on, and not necessarily a philosophical concern at all. Despite this, in theories of responsibility, concerns about character are almost always taken into account in some form or another. This is usually done in an arbitrary fashion, and rarely is its effect examined in sufficient depth. Whether this oversight is due to a need to distinguish one's position in the various classifications, or whether it is due to the so-called "is-ought gap", the options are brought up for discussion.

When the character-related instances are discussed in terms of new theory, without the original concept in place, the result can be quite convoluted, as is the case with Wallace and "blameworthiness inhibitors" (Wallace 1994). Inspired by Schoeman's question¹⁵ and by Wallace's wording, a unifying concept is referred to: *blame mitigation* is used here in order to describe the subject that is alternately dealt with in terms of character (as a philosophical concept) or excuses. Q2 is answered with the help of chapter 7's analysis. As a result, it is claimed that character or the concepts used in its place, such as the excuse, serve as necessary shorthand for the empirical data, or the aspect of *historical responsibility*¹⁶ needed in order to determine the fairness, blameworthiness and moral status of the wrong-doing agents in cases where there is possible variation between types of agent.

Strawson's choice to leave character-related concerns out of his theory of responsibility is recognized at this stage, his discussion of special considerations and pleas replacing the substantive content of the character-related discussion. Watson and Wallace expand the categories of these to excuses and exemptions. A revision of the concept of character is called for.

¹⁵ "How do our theories of character and our theories of excuses inform one another?" (Schoeman 1987, 8-9).

¹⁶ As discussed by Gary Watson in his 1987 article "Responsibility and the Limits of Evil" in Schoeman's *Responsibility, Character and the Emotions* (1987).

More specific character traits could help to formulate responsibility assessments with more descriptive power. As Doris (2002) points out, the old Aristotelian psychological concepts should not be reinstated, at least not without substantial revision. Clarity might come through for example the new findings of psychology and neuroscience, but the specific definition of these is left for another occasion. Further discussion is called for.

Structure

The concepts themselves and their examination will proceed as follows: Chapters 3 through 5 in Part I are focused on moral responsibility as a concept and how it is relevant for philosophical discussion. No claim is made that these chapters are comprehensive, nor that it would be possible to discuss everything related to moral responsibility. Rather, a selection of theories is discussed in order to sample the larger field of close or related theories; in other words, the study is presented in the form of a heuristic. Strawson's naturalistic, reactive sentiments-based conception of responsibility is selected as the point of comparison for the other theories. As the examination of the concepts becomes multidisciplinary, this first theory takes precedence, as the other theories are mainly discussed in terms of commentators from the naturalistic, Strawsonian discipline.¹⁷ The historical study of Aristotle is scholarly in that the relevant parts of his theory related to the modern concept of responsibility are compared with each other. This includes looking into Aristotle studies in ethics and political philosophy conducted in relatively recent years. The important ideas of Aristotle's theory will be examined in chapter 5 in order to describe an alternative conception of moral responsibility. Naturalistic interpretation of Aristotle is favored with a view to keeping the multi-disciplinary effort manageable.

Part II is structured around the analysis of the important concepts found in the theories described in Part I. The systematic discussion of the two key concepts, *character* and *excuse* will take place in chapters 6.1 and 6.2 respectively. Chapter six also includes two examples of extraordinary moral agents that have an effect on the responsibility assessments. Once the sections on the substance and the systematic exposition of the concepts in chapter 6 are done, the analysis based on Bernard Williams's explication of elements of responsibility is undertaken in chapter 7, wherein the comparative analysis between character and excuses in Strawsonian and Aristotelian theories of responsibility will be conducted. The argument leading to the thesis is summarized in chapter 8 along with discussion. Figure 1 below illustrates the structure of the dissertation:

¹⁷ A couple of examples: Aristotle's work is discussed here primarily through the naturalistic interpretation of Susan S. Meyer's *Aristotle on Moral Responsibility* (2011), which relies on Strawson. The most cited author on Hume is Paul Russell, whose *Freedom and Moral Sentiment* (1995) primarily compares Strawsonian and Humean theories of moral responsibility.

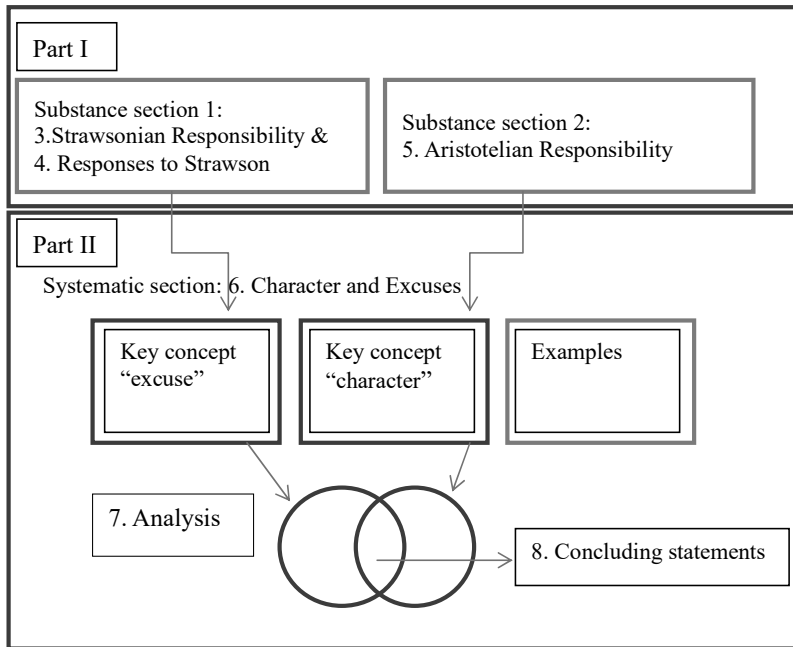


Figure 1 Structure of the dissertation.

2.2 LITERATURE REVIEW

Introductory texts on Moral Responsibility

In the “Oxford Bibliography on Moral Responsibility” (2010), Garrath Williams lists four general overviews on the topic, which are: A. Duff’s “Responsibility” (1998), A. Eshleman’s “Moral Responsibility” (2014/[2001]), M. McKenna’s “Compatibilism” (2004) and C. Kutz’s “Responsibility” (2002). Of these four, I have used Eshleman’s article, since that text’s emphases on Strawson and Aristotle are mirrored here. Strawson’s philosophy is further described in M. McKenna & P. Russell’s “Introduction” to *Free Will and Reactive Attitudes Perspectives on P.F. Strawson’s “Freedom and Resentment”* (2007).

Herbert Fingarette poses some seldom asked questions about the concept of responsibility in his article “Responsibility” (1967), declaring that the current understanding of the concept is not the only possible one, as discussed briefly in chapter 3.4.1. Even though Fingarette’s account lacks a specific definition of what responsibility actually is or what it should be, it holds up the example of an imperfect agent or even one with fundamentally different moralities (e.g., of a psychopath, which is an example that is discussed further on). Fingarette very straightforwardly offers an example that bears on the central issue of the most common theories of responsibility. Why are some agents excluded from even a basic assessment of whether or not they are morally responsible?¹⁸ Is responsibility a topic such that those incapable of understanding it are completely beyond its bounds? Fingarette leaves the issue open, but few after him have been as willing to question the basic premises of the topic. B. Williams, Smiley and Watson are examined here because of the diverging ideas that they bring to the discussion. In this chapter some of the most important alternative opinions on the nature of moral responsibility are looked at.

Strawsonian responsibility: “Freedom and Resentment” (2008/[1962])

The theory of reactive sentiments of Peter F. Strawson (1919 – 2006) is understood here to be something of a common denominator for writings on responsibility today at least in English (see Eshleman 2014, G. Williams 2010, ch.2).¹⁹ What this means is that when a student of philosophy gets

¹⁸ The debate surrounding the concepts of excuses and exemptions can be read as addressing these concerns (see ch. 6.2).

¹⁹ Ted Honderich, who is a key proponent of theories of determinism, gives the following description of Strawson’s importance:

The doyen of living English philosophers, by these reflections, took hold of and changed the outlook of a good many other philosophers, if not quite enough. He did so, essentially, by assuming that talk of freedom and responsibility is talk not of facts or truths, in a certain

interested in responsibility, the first work on the subject he or she most likely grabs is Peter Strawson's article "Freedom and Resentment" (2008/[1962] or alternately a work that it has directly influenced (and probably cited on the first page). This influential talk has functioned as the basis of the debates on responsibility within philosophy for the better part of a century now. Even when taking into account dissenting voices that were heard at the time — for example, Herbert Fingarette's psychologically minded essay, (see 4.3.1), which highlights the problem of applying philosophical responsibility to unconventional minds such as those of the psychopath — Strawson's success has been unprecedented.

His important "Freedom and Resentment" (2008/[1962], henceforth *FR*, examined in chapter 3.1), which Garrath Williams claims is the most influential article on moral responsibility, actually mentions concepts such as "moral responsibility", "blame" and "excuses" surprisingly sparsely (G. Williams 2010, 8). Not attempting to define moral responsibility explicitly, Strawson's initial goal was to *solve the problem of determinism* by reconciling the opposing viewpoints of compatibilists²⁰ (or optimists in Strawson's terms) and incompatibilists (pessimists). He was successful in that discussion of the concept of responsibility is no longer associated with them as often. It might have been even more successful if his article had actually brought the free will vs. determinism debate to a close. This, however, was not the case, as there still are incompatibilists doing a lot of work using the assumption of a merit-based conception of responsibility. There have also been new developments by theorists who have explored merit-based versions of compatibilism.²¹

Reactive emotions, like *resentment*, *indignation* and *guilt* are essential in determining these attitudes, which ultimately result in the act of holding someone morally responsible. The more important question should be about how the person is *held* morally responsible (Eshleman 2014), instead of asking whether responsibility assessments might depend on an actual theoretical judgment of one's *being responsible*. Additionally, Strawson

sense, but of our attitudes. His more explicit concern was to look again at the question of whether determinism and freedom are consistent with one another -- by shifting attention to certain personal rather than moral attitudes, first of all gratitude and resentment. In the end, he arrived at a kind of Compatibilist or, as he says, Optimist conclusion. That is no doubt a recommendation but not the largest recommendation of this splendidly rich piece of philosophy. (Ted Honderich on his website *The Determinism and Freedom Philosophy Website* <<http://www.ucl.ac.uk/~uctytho/dfwstrawson1.htm>>.)

See also Gary Watson's latest article "Peter Strawson on Responsibility and Sociality", in which he writes: "In the past twenty years, it has become routine to characterize responsibility in terms of the propriety conditions of what Strawson calls the reactive attitudes" (Watson 2014, 15).

²⁰ For more about the topic of compatibilism see McKenna's article on the subject at the *The Stanford Encyclopedia of Philosophy* –website (McKenna 2009).

²¹ Eshleman mentions Fischer & Ravizza 1998 and McKenna 2012 as sources in this matter.

attaches “*special considerations*” to resentment (a moral sentiment epitomizing directed blame), which can “modify or mollify” the sentiment. The “special considerations” are divided into two subcategories, the first of which could be described in Humean terms as “out-of-character”²² actions such as acting under the influence of excessive fatigue, alcohol, or some other temporary impairment to thought. The other category is more significant, containing both the undeveloped agency of children, as well as the “incapacitated agency” of mental illnesses and similar states. Albeit Strawson himself recognizes that his classification is “facile” (*FR*, 12), his followers have continued to elaborate on it. As a result, the group called “Strawsonian” here refers to Strawson as well as the group of followers he has inspired.²³ He concludes his essay by claiming that “It is a pity that talk of the moral sentiments has fallen out of favour. The phrase would be quite a good name for that *network of human attitudes* in acknowledging the character and place of which we find, I suggest, the only possibility of reconciling these disputants to each other and the facts.” (*Ibid.* 25-26. My emphasis.) As for the objections about different cultures, Strawson anticipated the critique that is also discussed in this study.

Strawson’s follower and critic Gary Watson explains that, while the connection between the reactive attitudes and the act of holding one responsible is not exclusive to Strawson’s theory, the original contribution of his work was that instead of being extraneous or a product of any metaphysical account of responsibility, the reactive attitudes themselves were constitutive of moral responsibility. The Strawsonian account of responsibility thus considers the reactive attitudes as the definitive signifiers of responsibility relations in all their forms. These reactions that occur as we regard ourselves and others responsible are inherent in our common practices as human beings. (See Watson 1987, 256-257.)

A significant virtue of the Strawsonian, revised compatibilist position is that it is an attempt to tone down the whole problem of determinism. Whether free will or determinism is true, at the end of the day it does not make any difference to how we judge behavior in our lives. Despite the opinions of whether Strawson’s refocus succeeded or not, the effort has had the positive effect of giving new wind to the philosophy of responsibility by

²² Hume considers character important in determining responsibility, writing that “We are never to consider any single action in our enquiries concerning the origin of morals; but only the quality or character from which the action proceeded. These alone are durable enough to affect our sentiments concerning the person” (Hume THN, as quoted by Russell 1995, 95). These “durable qualities of character” (see also THN, 398) have a major resemblance to Aristotle’s guiding virtues.

²³ The “Strawsonian” group refers here to Strawson himself as well as R. J. Wallace, who has developed Strawson’s theory of responsibility further (Wallace 1994). Other writers considered such, but referred to less here, include S. Darwall and S. Wolf (see Macnamara 2013, 141). G. Watson (see 1987) is also considered to belong to this group because of his extensive commentary on Strawson’s work.

giving the topic new tools to talk about the issue (incl. *reactive attitudes* and *sentiments* as well as the conceptual areas related to *excuses* and *exemptions*. etc.) The Strawsonian position on responsibility is not, however, by all accounts an ideal description of what responsibility is about. Therefore, as the common starting-point in discussion on responsibility it calls for arguments about the role of the concept, especially as it was never the writer's intention to give an exact definition of the concept of moral responsibility. Differing opinions about the concept have been advanced but, as a common point of reference, the Strawsonian position has remained influential, at times limiting the philosophical discussion of responsibility to a limited, act-based relationship between individuals. Yet it can be claimed that without the publication of *FR*, it is uncertain whether the subject of moral responsibility would be as discussed, or whether the naturalistic turn in ethics would have happened.

Aristotelian responsibility: Nicomachean Ethics and Politics

The works of Aristotle are considered to be the origin of the western philosophy of responsibility. As will be noted in case of the commentators (esp. Smiley 1992), the classical concept of responsibility differs from the modern one, and something has been lost in translation. The *Nicomachean Ethics* and *Politics* together form the corpus of Aristotle's political theory. They also contain writing on the topic of responsibility. Chapter III of the *Nicomachean Ethics* is particularly significant, because of its description of moral responsibility.²⁴ Aristotle's responsibility for action is looked at in chapter 5.1.1 of this thesis, while 5.1.2 focuses on responsibility for character. Chapter 5.1.3 considers an alternative perspective, by framing responsibility in terms of political as presented in the *Politics*. It could be argued that the entire history of the philosophy of responsibility is initiated in the *Nicomachean Ethics*, at least when the concept of responsibility is based on the dual concepts of praise and blame. These two classics by Aristotle will be used here ultimately to outline his concept of character as part of his theory of responsibility, which is the topic of chapter 6.1.. The *Eudemian Ethics* and *Rhetoric* are referred to to a lesser extent. The same goes for the commentaries mentioned by G. Williams (2010), which are S. Broadie (1991) & S.S. Meyer (2011). Broadie and Meyer are referred to a lot in terms of ethics. Additional commentaries on the political side of Aristotle are provided by D. Winthrop (1975) and M. P. Nichols (1992).

Even though the work of Aristotle predates the modern theory of Strawson by quite a stretch, the different theories will appear in an inverse chronological order so as to give more weight to the current philosophical theory that receives the most attention as well as grounding the multi-

²⁴ Aristotle's responsibility is challenging to define, as he did not use a directly translatable term. Chapter 5 contains an account of the subject.

disciplinary study in the broadly naturalistic school of ethics. This is possible through various philosophers who study the Aristotelian subject in terms of the Strawsonian theory of responsibility (especially Meyer 2011; see also Irwin 1980).

Aristotle and his writings on responsibility are of great significance to this thesis. A previously published, peer-reviewed article (and a translation presented at the XXIII World Conference of Philosophy) written by the author is used as a basis for chapter 5.

Responses to Strawson

R. Jay Wallace's version of Strawsonian moral responsibility will be discussed as an extension to the theory with a normative goal. As these are distinct parts of Wallace's own theory of responsibility, they are looked at in chapters 4.1.2 and 4.1.3 respectively. These ideas are compared to the character-based approaches in chapter 6. Regarding the concept of responsibility, both Strawson and Wallace are considered as members of the Strawsonian theories, which is subject to criticism from alternative viewpoints.

Since the debate on the concept of responsibility has been going on for the more than five decades since *FR*, it is naturally understandable that the topic takes new forms in order to keep the interest of scholars. Thus the philosophical discussion has focused in recent years on the concept of blame, which is a significant component of every theory of responsibility. The work of Coates & Tognazzini (2013 & 2014) is used primarily to explore this topic. The group of philosophers focusing on the subject are those who have been of recent interest for the author but, at the time of completing this thesis, they have been deemed too much to fully accommodate. These topics will very likely influence any further forays into the subject of responsibility.²⁵

General Criticisms on Moral Responsibility

While Michael McKenna & Paul Russell's *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment"* (2007) is the first critical source that is used here to pinpoint the most obvious difficulty with Strawson's theory, Herbert Fingarette's "Responsibility" (1966) works as a contrasting example of how the concept could be defined differently than is currently considered customary. Continuing the critical section, Marion Smiley effortlessly demonstrates in her *Moral Responsibility and the Boundaries of Community* (1992) that responsibility has not always been as narrowly defined; that there are problems with this definition, and that what

²⁵ Following Victoria McGeer's list: Nomy Arpaly, Pamela Hieronymi, Gideon Rosen, T. M. Scanlon, George Sher, Angela Smith, Gary Watson and Susan Wolf, among others, have written on the topic of blame (McGeer 2013, 162-163.)

is left out of the modern definition contains some interesting things worth reconsidering. Finally, Bernard Williams argues that there can never even be a proper, comprehensive definition of responsibility. This is because all attempts to formulate a theory of responsibility must incorporate of four elements of the concept (1. cause, 2. intention 3. state and 4. response), which are often given different weight or left ill-defined. In the 1997 article, these are stated more explicitly, are of tremendous help in clarifying the concept as such, and are scrutinized in chapter 3.4.3. The elements are derived through a historical analysis in his *Shame and Necessity* (B. Williams 1993), where Williams examines the case of responsibility in ancient Greece. Aristotle's theory appears in that analysis. The classification might oversimplify the related issues, but it is a very apt one, and it is important that Williams pinpoints the element of "state"²⁶, which is insufficiently addressed in most accounts of responsibility in contemporary philosophy. (See ch.5.) The classification is further utilized in the analysis section in chapter 6.

Direct Criticism on Strawson

Gary Watson's "Responsibility and the Limits of Evil" (1987) published in Ferdinand Schoeman's equally interesting book (1987)²⁷ is very important in introducing the division into excuses and exemptions, which Wallace adopts as a basis for his typology of excuses and exemptions (Watson 1987, 259-261). This subject is discussed extensively in chapter 6.2. Others of Strawson's critics referred to here include Fischer & Ravizza (1998) and Paul Russell²⁸ (2013).

²⁶ Bernard Williams uses the element of state to denote mental states that are affecting the act and that are relevant in terms of moral responsibility for that act.

²⁷ Ferdinand Schoeman's 1987 anthology *Responsibility, Character and the Emotions* is a very important book not only because of its inclusion of Gary Watson's article, which was the first to distinguish the concepts of excuses and exemptions, but the book is also worth a read for the introduction itself. Schoemann posits some very interesting questions, one of which has influenced the topic of this dissertation itself, Schoeman's question "How do our theories of character and our theories of excuses inform one another?" being an important inspiration (Schoeman 1987, 8-9). *Responsibility, Character and the Emotions* is from its outset very close in its goals to this work and is thus very useful and interesting. The only disadvantage of this collection is that it dates back to 1987, so that most of the commentaries that form the structure for this work did not exist then.

²⁸ Paul Russell is primarily an accomplished Hume scholar whose works have compared Hume's and Strawson's ideas. Most relevant of these works is the impressive *Freedom and Moral Sentiment* (1995), which discusses both the similarities of Hume's and Strawson's theories as well as the specific writings on responsibility by Hume. Russell objects to Hume on character, labeling Hume's musings on character in the context of responsibility essentially a mistake. While Hume is not under direct investigation here, overall Russell's book is a considerable asset in charting Hume's influence on Strawson's philosophy.

Commentators on Aristotle

Since introductions to Aristotle's thought are numerous, the most recent of these, by Susanne Bobzien (2014), is mentioned here. This article describes the area of moral responsibility within Aristotle's texts quite clearly. In order to make the broad multi-disciplinary study possible in which current theories are contrasted with their historical counterparts, priorities need to be clear. Therefore most emphasis is given to the explicitly Strawsonian, naturalistic interpretation by Susan S. Meyer in her *Aristotle on Moral Responsibility* (2011). The relevant debate on Aristotle's moral responsibility, discussed in chapter 5.2, revolves around the proposed effect that moral character has on moral responsibility. Interlocutors of this discussion who are mentioned here include Terence H. Irwin (1980), Jean Roberts (1989), Sarah Broadie (1991) and Randall Curren (1989).

Specific sources on individual concepts

As was the case with *blame* and the works of Justin Coates & Neal A. Tognazzini, several sources have been consulted, especially in part II, because of their contribution toward particular individual concepts. Marcia Homiak (2015) is discussed in chapter 6.1 as a primer for study on *moral character*. Likewise, the critical notes by John Doris (2002) are looked at as being the most pressing criticism on the subject. Although Doris hails from social psychology, his criticism is so encompassing that philosophers have had to answer it. Mark Alfano (2013a) has grouped the answers to the "attack on character" topic in the three categories of *dodge*, *retreat* and *counterattack*, where the "dodgers" deny the threat that situationism poses for virtue ethics;²⁹ the "retreaters" claim that even though situationism is correct, there is a weakened core to the virtue ethics that may be salvaged; as well as the position of the "counterattackers", according to which the situationist challenge fails outright (Alfano 2013, 244-246).

Similarly *excuse* is looked at in chapter 6.2 as its own topic. Starting with J. L. Austin's ideas in his article (1956/7), the description of the concept of excuse is clarified by looking at a couple of more modern texts on the subject. The first of these is by Strawson himself, who is succeeded by his commentators Watson (1987) & Wallace (1994). Holly M. Smith's alternative

²⁹ For example, Christian B. Miller has recently done a lot of work in defense of the philosophical notion of virtue, relying on original Aristotelian ideas (see Miller 2013 & Miller, Furr, Knobel & Fleeson (eds.) 2015).

PART I: TWO THEORIES OF RESPONSIBILITY

3 STRAWSONIAN RESPONSIBILITY

What is compelling about responsibility is that there has never been a single unified, unanimously accepted understanding about it. The concept has been a subject of philosophical inquiry since the very beginning of Western philosophy and it has remained a regular staple of the field to the extent that it has gained great diversity as a subject. It is not uncommon that the various schools of thought, making the effort to make the concept clear, are baffled about whether they are talking about the same subject at all.

The word “responsibility” entered the English language philosophy in the 17th century from the arena of political debate.³⁰ Relying only on its origin would produce a very limited definition of the concept, for the area of research that the word “responsibility” signifies has been a subject of vibrant debate now for millennia. The disputes have returned from time to time to the question of blame and causal responsibility, as well as moral agency that are now commonly understood as being part of the responsibility debate (see G. Williams, 2010). To accommodate all theories and opinions and to be able to address them here would be impossible. The works concerning the subject are now so numerous that this study will be offered in the form of a heuristic, with two major, significant approaches to the subject that are taken as representing the subject area. These approaches are: (1.) the reactive sentiments based, Strawsonian approach, discussed in this chapter, and (2.) Aristotle’s responsibility according to contemporary scholars, which is looked at in chapter 5. These two approaches will then be subjected to comparative analysis, focusing on moral responsibility, blame and its mitigation through the concepts of character and excuses, examples and criticism. The analysis concentrates on the different concepts of character and excuse. This will be undertaken in the second part “Character and Excuses in Theories of Responsibility”. For now, the focus will be on the more general presentation of the subject.

Responsibility itself has proven to be an elusive concept to define. Surprisingly (or perhaps precisely because of this fact) fully formulated

³⁰ Garrath Williams writes in his “Responsibility as a Virtue” relying on the *Oxford English Dictionary*: “Its most important original use was in political thought and debate; for instance, in the Federalist Papers (Hamilton et al. 1787, no. 63) and Edmund Burke (1796, p. 508)” (G. Williams 2008, 457). However, Marion Smiley notes in *Moral Responsibility and the Boundaries of Community* (1992) that “Albert Jonsen writes in *Responsibility in Modern Religious Ethics* (Washington, 1968) that the term appeared somewhat earlier in French, most notably in Pascal’s *Lettres Provinciales* (1656), although Jonsen fails to point out that Pascal’s words are ‘responsibility and blame’, rather than ‘moral responsibility’.” (Smiley 1992, 35-36, n9.) The adjective seems to date back even further.

theories of responsibility³¹ can be difficult to identify as such.³² Even as moral responsibility is the main concept that frames the thesis here, the notion of moral responsibility is formidably diverse, and in order to focus the issue some cuts and further explication of concepts have been made. *Responsibility* at the most general level can mean, among other things, moral responsibility, legal responsibility, political responsibility, the property of being responsible (a virtue or an excellence), etc. *Moral responsibility*, in this case the category from these options, is that which gets the most attention here. Moral responsibility is closely related to such concepts as moral agency (which is a common requirement for being responsible), accountability (mostly through blame) and blameworthiness (an element of moral responsibility distinct from causal responsibility, which will be examined below), among other things. Moral responsibility, following Bernard Williams's writings, is examined through its elements of cause, intention, state and response. The last of these is a common feature in theories of responsibility, and has been divided into praise and blame since Aristotle. Since blame is a popular topic for study today, it will be featured with the two main theories of responsibility considered in this work. Finally, the different types of excuse in the case of Strawsonian responsibility and character in Aristotle's counterpart are discussed as forms of blame mitigation.

As a point of departure, Andrew Eshleman's article on "Moral Responsibility" (2014/[2001]) in the *Stanford Encyclopedia of Philosophy* presents a selection of criteria, according to which theories of responsibility can be identified.³³ It provides an account of theories of *moral responsibility*,

³¹ The word "theory" here is used in the broadest possible sense. "Accounts of responsibility" is used here interchangeably with the previous term. The phrase "conception of responsibility" is used below to refer to the identical subject matter with less need for validation. Accuracy typical of the natural sciences is not the aim here, as Strawson's example is followed here in: "my language, like that of commonplaces generally, will be quite unscientific and imprecise" (*FR*, 5). Further, even with Eshleman's criteria, locating or pinning down the exact theory in these cases would be difficult. Criterion 1 in Aristotle's case is arguable. *Aitia*, meaning *cause*, is probably the Aristotelean concept that comes closest to a translation of the modern word. A lot of his responsibility-related content is discussed on the other hand through *telos*. These are some of the considerations that have been taken into account throughout this study. (See Eshleman 2014/[2001].)

³² The same seems to be true in cases of theories of blame. (See Coates & Tognazzini 2014.)

³³ Eshleman's *Stanford Encyclopedia* overview of "Moral Responsibility" (2014/[2001]) is structured around *FR*, the article by Peter F. Strawson in question, which has had a tremendous influence on the subject. Published on one of today's premiere internet sources, in the *Stanford Encyclopedia of Philosophy*, Andrew Eshleman's "Moral Responsibility" article is a fine introduction to the subject, in which the Strawsonian, or the reactive sentiments approach to responsibility is discussed thoroughly (see ch. 3.1). Historically speaking, Strawson's article, according to Eshleman, has enabled the continuation of the debate about the concept of responsibility, as it has since the defined the limits of most work on the subject.

focusing on Peter F. Strawson's reactive sentiments based view of responsibility. Eshleman outlines requirements for theories of moral responsibility that can cover this sufficiently to be called such. He includes the following elements in his list of what theories of responsibility should contain: first an *idea or a concept of moral responsibility* itself; second, *the criteria of moral agency* or what is required from a moral agent; third, the *conditions of application of moral responsibility* and, finally, the possible *objects of responsibility ascriptions*, including actions, omissions consequences and character traits. (Eshleman 2014/[2001].)

To begin with, Eshleman discusses two different theorists of responsibility in his article, Aristotle and Strawson; neither of whom explicitly built a definition of responsibility as the word is understood now, but who nevertheless have both become exemplary for the following discussion. His focus is on the distinction between *compatibilist consequentialism* and *merit-based incompatibilism*.

Discussing *FR*, Strawson's "landmark essay" further, Eshleman describes how Strawson sought to end the disagreement between the consequentialist compatibilists (or optimists, in Strawson's words) and the incompatibilists (pessimists), who defined responsibility as being merit-based. The other division in Eshleman, along with the compatibilist/incompatibilist one, is between the *merit-based view*, which considers praise and blame appropriate according to the deserts of the agent's actions, and the *consequentialist view* that states that praise and blame are appropriate if and only if the response instigates a desired change in the behavior of the agent. (Eshleman 2014/[2001].) Following this elaboration, Eshleman concludes that "Strawson's [theory of moral responsibility] is a merit-based form of compatibilism." (Ibid.) As Eshleman writes, however, his position is compatibilist only by default, Strawson's ultimate intention being to leave the free will vs. determinism debate behind.³⁴

In addition to the distinctions of responsibility utilized by Eshleman, there are many other possible conceptual distinctions that can help to situate the relevant content of theories as well as decipher one's views on responsibility. The following distinctions are discussed throughout this work:

- Optimism and pessimism

Optimism and pessimism in this context correspond to Strawson's words for compatibilism and incompatibilism respectively. For the first half of the 20th

³⁴ While the division between consequentialist compatibilists and merit-based incompatibilists is interesting and works well as the backdrop for Eshleman's argument, it is not pursued further here. The same goes for Strawson's optimist/pessimist pair of descriptors. The classifications make sense in the context of the problem of determinism, but they can be misleading otherwise. Incompatibilism, for instance, contains both hard determinists and metaphysical libertarians. While it has been noted, the free will vs. determinism debate is not the main concern here.

century, the complete Anglophone school of ethics seemed basically to be mired in unresolvable debates. The really hard problem, however, that endures to this day, is the problem of determinism. If determinism were accepted, moral responsibility would be outright impossible. The proponents of the principle of determinism, *hard determinists*, have sought to reformulate ethical practices from the ground up in order to mirror this principle, while those who deny determinism and emphasize free will (or *libertarians*) have frequently run into trouble in basing their theories satisfactorily in the face of contradictory evidence. In the debate on responsibility, the camps are usually divided into compatibilist and incompatibilist. This notion of compatibilism refers to the possibility of reconciliation between determinism and morality, or essentially the answer to the question of how morality possible is if reality is causally determined. This distinction is used here only in connection with Strawson's article. As the emotional reaction based classification has not caught on since, and since writers still refer to the concept pair as compatibilist/incompatibilism, the latter form is returned to once the section on *FR* is at an end.

- Causal responsibility and blameworthiness

According to Smiley, the ideas of *causal responsibility* and *blameworthiness* overlap in the aggregate concept of moral responsibility (1992).³⁵ Causal responsibility is what happens when someone causes something by acting. In the case of pure causal responsibility, it is irrelevant what the intentions of the actor or the circumstances were.³⁶ Since Aristotle, blameworthiness has been used as a requirement for the more imprecise concept of responsibility. If one is blameworthy, one can be considered morally responsible. If one is not, being morally responsible is also out of the question. As a more pragmatic clarification, Smiley's point about the conflation of the concept of

³⁵ Marion Smiley's 1992 book *Moral Responsibility and the Boundaries of Community* is an important departure from the mainstream conceptualizations of responsibility as it separates three varieties of responsibility (classical, Christian and modern) on a historical scale. Additionally Smiley emphasizes the issue about the philosophy of responsibility conflating questions of causal responsibility and blameworthiness. These distinctions are examined in this thesis starting at chapter 4.3.2. According to Smiley there are three important versions of responsibility throughout the history of Western philosophy (Smiley 1992). The modern concept of responsibility is distinct from two other historical concepts: the classical and the Christian. The modern concept is basically the same as the latter but, according to Smiley, merely removed to a secular context. Smiley describes the modern concept as not much more than a "secularized version of Christian sinfulness" (Smiley 1992, 73, note 4).

³⁶ Coates & Tognazzini (2014) also use the synonymous term of *explanatory responsibility* with causal responsibility. Their article "Blame" notes that when discussing moral responsibility most philosophers ignore causal responsibility altogether, since it does not necessarily take a moral agent to cause something to happen. (Coates & Tognazzini, 2014.)

moral responsibility is supported here. Blameworthiness is a key concept that is used universally here to denote actual moral responsibility. (See also Praise & Blame below.)

- Being responsible and holding responsible

A sometimes unnoticed mix-up is pointed out by Fischer & Ravizza (1998) and even more explicitly by Angela M. Smith (2007) is that between *being* responsible and *holding* responsible.³⁷ Most of the writings discussed below deal with the “holding responsible” side, because determining whether one “is” actually responsible is difficult if the context is not in a set, legal environment. The remaining area of the “being responsible” side, while not actively dismissed, reaches toward the metaphysical, which for obvious reasons is an aspect not pursued in this limited-length dissertation. Strawson is said to distinguish between the two insufficiently, while Wallace explicitly leans toward the holding side.

- Cause, intention, state and response (four elements of responsibility, according to Bernard Williams)

Bernard Williams’s (1993 & 1997) explicates the elements present in full theories of responsibility.³⁸ These are *cause*, *intention*, *state* and *response*. More detail on this can be found in his texts (see also chapter 4.3.3). His distinction of four elements of responsibility is an important tool that is used in analyzing the current form of the concept of responsibility (chapter 7). The final element, response, is discussed in all of the theories discussed here in terms of praise and blame, the latter having become the topic of recent interest.

³⁷ Angela M. Smith provides one of the most recent, useful classifications of responsibility in her “On Being and Holding Responsible” (2007). For example, Wallace devotes his research explicitly to the “holding responsible” side of the split. How this is problematic is examined in chapter 4.1.4. Her article “Moral Blame and Moral Protest” in Coates & Tognazzini (2013) is also noteworthy for offering a defense of an account of blame, which continues Scanlon’s and Sher’s work on this concept (see Smith 2013, 47).

³⁸ The work of Bernard Williams throughout the last decades of the 20th century is without peer. He devoted some attention to the concept of responsibility in the 1993 *Shame and Necessity* as well as the article “Moral Responsibility and Political Freedom” from 1997, in which he discusses the differences between the concepts of responsibility held by the ancient Greek philosophers and their current counterparts.

- Praise and blame

Present in theories of responsibility since Aristotle's *Nicomachean Ethics*, *praise* and *blame* are the distinct forms of reacting to morally relevant actions, be it wrong-doing or right. The interest of the philosophers seems lately to have gravitated toward the latter at the expense of the former. The most recent discussion on blame is briefly summarized in chapter 4.2.

- Responsibility for action and responsibility for character

A specific distinction helping to point out the special role for his addendum is the theory of *responsibility for action* presented in *Nicomachean Ethics* III.1. Passages on the distinct topic of *responsibility for character* are found in book III.5 of *NE*. The examination of these concepts is the subject of chapter 5 of this dissertation and the curious role of the latter idea is elaborated on in the discussion section of that chapter. The more concrete concept of political responsibility is part of the continuation of Aristotle's writings on responsibility. While mentioned, this specific social idea is not in the main interest here.

- Accountability and attributability

Gary Watson's distinction between *accountability* and *attributability*³⁹ is an interesting division. Watson expresses the idea that focusing only on one side of the responsibility "coin" leaves out something important, since being accountable is more retrospective⁴⁰ by nature but, viewed without attributability, which is a prospective form of responsibility, might produce overly mechanical systems of fact-checking. Eshleemann states that the clearest example of the accountability-based accounts of responsibility are the Strawsonian theories based on reactive attitudes (incl. Wallace 1994; Watson 1996; Fischer & Ravizza 1998; and Darwall 2006), Accounts making no reference to the reactive attitudes however include the "ledger" views of

³⁹ Creation of this division is credited to Gary Watson (1996, 271-272, 273-274, also reprinted in Watson 2004), although similar divisions have been made as early as Frankena (1973), who described three possible types of responsibility: 1. X could have responsible moral character (character-based responsibility, responsibility as a virtue); 2. X could be responsible for Y, where Y is a past crime (retrospective accountability); and 3. X is responsible for Y, where Y is something still to be done (prospective attributability) (Frankena 1973, 71). Also of note is Watson's proposed hybrid concept of *answerability*, which would be quite close in meaning to the Finnish language equivalent, "vastuullisuus".

⁴⁰ For illustration of the temporal distinctions between responsibility, *retrospective responsibility* and *prospective responsibility*, see Garrath Williams's internet article "Responsibility" (G. Williams 2009).

moral responsibility, which consider responsibility in terms of credit and debit on a metaphorical ledger of conduct associated with each moral agent. (See Feinberg's work as an example.) (Eshleman 2014/[2001].) On the other hand, looking just at the attributability side is not enough either. According to David O. Brink and Dana K. Nelkin "[a]ttributability is necessary but not sufficient for accountability" (Brink & Nelkin 2013, 286).⁴¹ This is compounded by Eshleman's note stating the questionable status of attributability as a form of responsibility (Eshleman 2014).

As Strawson's influential contribution depends solely on the relatively short article *FR*, originally delivered as a lecture to the British Academy in 1962, the details of his theory have proved to be contested (see Strawson 2008, xii). After all, the article itself never aimed to explicitly define responsibility. As the debate on the concept has since advanced to refine its more minute details, an attempt to define Strawsonian responsibility based on the article alone, would not produce an accurate description of the subject. As support, the popular, "contemporary Strawsonian" R. Jay Wallace is looked at in chapter 4.1, as his book *Responsibility and the Moral Sentiments* (1994) provides perhaps the most fleshed-out theory, which is mostly expanded directly from the Strawsonian position outlined in the article. Wallace's *Responsibility and the Moral Sentiments* has in turn become a staple of moral theory and is often returned to when the Strawsonian position is adduced. This is also the case in the currently popular academic discussion on the concept of blame. These discussions, and how they frame the Strawsonian position in relation to the more recent attempts at defining blame are examined as the final viewpoint on the Strawsonian concept of responsibility (4.2).⁴²

Selected criticisms of the Strawsonian conception will conclude this chapter (4.3). These criticisms will highlight the delineation of the concept of moral responsibility relative to the more general topic of responsibility (4.3.1); offer additional distinctions that help locate the important issues

⁴¹ David O. Brink & Dana K. Nelkin cover a lot of common ground that is looked at here in their "Fairness and Architecture of Responsibility" in Shoemaker (ed.) 2013. What is interesting about their article is the claim that Strawson (2008/[1962]) can be interpreted in two ways: first in a *response-dependent* way and second in a *realist* way. The writers consider the response-independent interpretation of Strawson, which is a different goal than what is pursued here. I argue that the response (mostly blame) is integral to the Strawsonian conception of responsibility, i.e., the "response-dependent" interpretation is followed. Other interesting contrasts that Brink and Nelkin make include a note on the justifications and excuses in the context of moral responsibility. It is implied that this concept-pair originates from the domain of criminal law. Generally Brink and Nelkin's article is situated between moral responsibility and criminal law in terms of discipline, and would be a very useful bridge between the subjects (see Brink & Nelkin 2013, 285-90).

⁴² The Strawsonian conception of responsibility discussed here is based on its subsequent development by works including (but not limited to) Watson 1987; Russell 1995; Fischer and Ravizza 1998; Wallace 1994; McKenna & Russell, (eds.) 2007; and Russell 2013.

related to the concept as well as contrast the historical alternative conceptualizations of responsibility with the current one (4.3.2); question the project of attempting to force the subject area under one concept in the first place, as well as providing a framework for elements of all theories of responsibility (4.3.3); and finally question the stances that the Strawsonian position takes toward the role of historical responsibility as well as the relatedness of moral responsibility and blame (4.3.4).

Before proceeding to these specific areas, the original article, "Freedom and Resentment" is discussed next. The problem that Strawson set out to solve is explained first in 3.1.1, the reactive sentiments and blame are discussed second in 3.1.2, special considerations and pleas are the subject of chapter 3.1.3 and finally Strawson's theory's social dimension is described in chapter 3.1.4.

3.1 RESPONSIBILITY IN “FREEDOM AND RESENTMENT”

Paul Snowden describes the naturalistic turn achieved by Strawson’s essay in the following way in his *Stanford Encyclopedia of Philosophy* article (2009) on Strawson’s philosophy:

Strawson's purpose is to dissolve the so called problem of determinism and responsibility. His argument is that our 'reactive attitudes' towards others and ourselves, such attitudes as gratitude, anger, sympathy and resentment, are natural and irrevocable. Their presence, therefore, needs no abstract entitlement from philosophy, which is simply irrelevant to their existence or justification. There cannot be abstract a priori principles locating general metaphysical conditions for such attitudes. His claim is that our practice of holding ourselves and each other responsible for actions is similarly natural and not dependent on general metaphysical requirements. Between determinism and responsibility there can be no conflict. (Snowdon, 2009).

Peter Strawson’s philosophy is considered to reside within the tradition of broadly analytical ethics, as well as the school of ordinary language philosophy (see Parker-Ryan 2012). And while his contribution to moral philosophy is small in quantity, his influence is greater than most others. His essay “Freedom and Resentment” (Strawson 2008/[1962], abbr. *FR*), which can be read as a response to emotivists, utilitarian instrumentalists (such as Nowell-Smith 1948 & 1960) and moral skeptics, was originally presented as a lecture to the British Academy in 1962. Strawson tries to address an impasse that moral philosophy in general had arrived at. The debate in question is between the groups of philosophers now known as compatibilists and incompatibilists, who held opposing views in the determinism vs. free will debate. Strawson’s essay is constitutive to the position that accepts both, that the natural world, which we inhabit is deterministic by nature and that moral responsibility is possible to attain for human beings, i.e., compatibilism. As stated in Snowden’s foreword of the collection: “It would [...] be highly inconvenient should it be true that falling under the idea of responsibility requires something that we lack” (*FR*, xiv).

Strawson's most influential contributions to ethics within *FR* are, first his attempt to resolve the free will vs. determinism impasse by a redefinition of the compatibilist position, and second, related to the first, his grounding of responsibility directly on a naturalistic view of human psychology by attaching more relevance to emotions. What was more important was not whether moral statements have truth-values, e.g., that “Stealing is bad” is true or false, but reactive sentiments, such as resentment and gratitude, praise and blame.

Strawson's article is often mentioned by scholars on responsibility, but his ideas are seldom examined in great detail.⁴³ In any case, he and his article seem to be the person and the text that are associated most often with philosophical moral responsibility. While this is a claim that needs to be made with caution, there are entries supporting it: Garrath Williams calls *FR* "The single most influential essay in the literature on moral responsibility".⁴⁴ Williams describes the function of the article as one moving the debate on responsibility away from whether we are discussing moral facts or values to the very personal moral attitudes such as resentment or gratitude. (G. Williams 2010.)

⁴³ Referring directly to *FR* risks the misconception that the article could contain a definitive account or theory of the concept of moral responsibility. It of course does not attempt to be comprehensive in this sense; for example, in the passage above, Strawson states that the tone of his text will be "unscientific and imprecise" in the early pages of the article (*FR*, 5).

⁴⁴ *FR* is widely cited in the literature referred to here. Examples include: Wallace (1994, throughout), P. Russell (1995, pp.5, 75, [...] chapter 5), Smiley (1992, pp. 15, 77, 79, 242-243, 253), Doris (2002, p. 129, incl. a chapter on responsibility), Waller (2011, pp. 1, 15, 180, 190, [...]), G. Williams (2008, p. 461), (2010, p. 8.). On the other hand, Aristotle scholars refer to it infrequently, with Meyer (2011) as a notable exception. She has written explicitly from the perspective of broadly naturalistic ethics and is thus influenced by the article (see Meyer 2011, 161).

3.1.1 THE PROBLEM OF DETERMINISM

Strawson's renowned article does not set out just to define a concept of moral responsibility explicitly. At the beginning of *FR* he wonders what the true concern of the thesis of determinism might be. He admits up front that he does not have a clear idea about what determinism actually is.⁴⁵ He continues however: "perhaps we shall see that the question can be answered *without* knowing exactly what the thesis of determinism is" (*FR*, 11). It is a comment on the discussion of the determinism vs. free will debate and how it relates to moral responsibility in general at a time when it was thought that conceptual analysis could solve the apparent contradictions between the phrases in the supposed ordinary language (see *ibid.*, 1).

The reason for Strawson's success is in part to do with the free will problem or, more accurately, that Strawson's article provides the keys to circumvent it. Before Strawson, questions about responsibility were focused on the sort of fact-value discussions that were favored by the logical positivists for one. The proponents of determinism focus on the seemingly impossible nature of responsibility assessments as *facts* as these were considered to be moral statements with values at their heart (moral statements, according to the logical positivists, who were emotivists by default, do not have logically determinable truth-values).

Strawson follows this initial formulation with his classification of the disputants into optimists and pessimists depending on the attitudes that the writers hold toward the acknowledged possibility of the veracity of the determinism thesis (*ibid.* 2-3). The issue between the "optimist" and "pessimist" stances is the most striking, as Strawson writes, in cases of retribution as well as moral condemnation and approval. For the pessimist in particular, the schism between the determinists and libertarians is crucial. While the division concerns the age-old question about the origin of actions, he diverts attention toward emotions such as "resentment, forgiveness, [and] love".⁴⁶ (*Ibid.* 5.) According to Strawson it's not important to discuss whether

⁴⁵ A possible distinction that Strawson alludes to could be that between physical determination and theological determination. Beginning from Campbell 1951 and onwards, a growing number of commentators feel that the whole debate on free will and determinism is misleading, and that it really does not contribute at all to the philosophical investigation into the nature of responsibility (see Smiley 1992). The theme is however so pervasive that it has to be acknowledged in connection with responsibility and it should be noted, especially if a writer has a specific stance on the issue. Therefore, when relevant and clear, the position of the author on the compatibilism-incompatibilism divide will be stated, but a more extensive analysis of what the division implies is left to another occasion.

⁴⁶ For examples of philosophy of emotions from the pre-Strawson era, see Schlick 1966, Nowell-Smith 1948, Campbell, C.A. 1951 & Austin 1956-7. Strawson also refers to La Rochefoucauld, whose ideas are discussed in Christopher Tilmouth's *Passion's Triumph over Reason*, which is a very interesting intellectual history on the rise of the philosophical interest in the emotions. (Tilmouth 2010, 273, 300, 304-305.)

the statements above are facts or not, but rather what motivates the claims based on the sentiments that we have in connection with responsibility assessments. These motivations consist, according to Strawson, of emotions such as resentment, anger, gratitude and praise, etc. He explains that these attitudes are crucial to our capacity to distinguish between those who qualify as responsible persons and those who do not, and are non-responsible agents.

Michael McKenna and Paul Russell's book *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment"* (2007) contains articles related to the influential essay by Strawson.⁴⁷ The introductory essay describes "Freedom and Resentment", originally published in 1962, as being:

among the most important and influential of the contributions produced during this period. [- -] On one side, it must be viewed from the perspective of the "classical" free will debate as it was generally understood around the middle of the twentieth century. On the other side, we need to consider "Freedom and Resentment" in terms of the critical responses and debates that it has generated. (McKenna & Russell 2007, 1.)

Strawson's goal was ambitious in aiming to resolve the problem of determinism. The classical free will debate, he argues, can be described in terms of this problem. Determinism in this context is usually as thought about as follows: "everything that happens in the world—including all human thought and action—is subject to causal laws and this involves the necessitation of effects by antecedent causal conditions." (McKenna & Russell 2007, 1.)

If this is true, and all of our actions are necessitated causally, free will and moral responsibility appear to be impossible. The trouble is that free will and moral responsibility *require* causally necessitated actions. This leads to the situation in which no matter what our interpretation of the condition of determinism is – true or false – the ideas of free will and moral responsibility become logically untenable. (McKenna & Russell 2007, 2.)

The conflicting state for free will and moral responsibility in connection with determinism leads to the skeptical conclusion that there is no room for this concept. The two rival points of view are borne out of this situation. The classical compatibilist position and the variations of incompatibilism refer to this dilemma: compatibilism states that causal necessitation can co-exist with determinism. Classical incompatibilists vary in the strategy they use to

⁴⁷ In addition to these, their co-authored introduction is one of the best introductions to Strawson's best-known article as well as his thought. The essay is divided into three parts titled: "I. The Classical Debate and the Dilemma of Determinism", "II. Strawson's "Reconciling Project" and the Naturalistic Turn" and "III. Critical Themes Concerning "Freedom and Resentment". This structure is also followed loosely here.

advance their arguments.⁴⁸ The idea of forward-looking, utilitarian moral responsibility is dismissed by the incompatibilist, but accepted by the compatibilist. Campbell, for example, who McKenna & Russell refer to, noted that while our conduct can be altered by means of punishments and rewards, responsibility itself is not limited only to these (see Campbell, C.A. 1951). This is because animals and children can be observed to react similarly to punishment and reward, without them qualifying as moral agents.

The incompatibilists take hold of this, as they criticize the compatibilists for not having a sufficiently “*deep*” account of moral responsibility⁴⁹. Strawson adds the considerations of desert and identification of the “will with the act”, and the observation that the emphasis expressed in these viewpoints can lead to attempts to disprove the thesis of determinism. The

⁴⁸ As a group, incompatibilism includes the polar opposites of the metaphysical *libertarians* as well as *hard determinists*. Cross-referencing the stances on determinism and free will reveals a possible explanation for this :

	Free will is true	Free will is false
Determinism is true	Soft determinism / compatibilism	Hard determinism
Determinism is false	Libertarianism	Soft incompatibilism

The debate has mostly to do with underlying worldviews of the authors and mostly these do not directly influence the actual events that need moral consideration. Thus, the incompatibilists are a group consisting of both libertarians and hard determinists: Metaphysical libertarians claim that determinism is false, and that free will is an actuality and is at the heart of morality. Thus any moral decisions are ultimately connected to the agent or the person who decides to act. The libertarian stance runs into trouble when confronted by empirical evidence against the capacity for these rational decisions. Hard determinists on the other hand adhere to determinism and absolve all the potential for moral responsibility of an individual. According to libertarians, moral responsibility requires that the choice between alternate courses of action be made by the agent herself. This would mean the falsity of the condition of determinism. On the other hand, the hard determinists maintain that the reality is completely determined, and as such even the practice of blaming is questionable on the grounds that the agents are unable to affect anything (See McKenna & Russell 2007, 3 & Waller 2011). Most people find the resulting demands made by the hard determinists, such as abolishing moral responsibility, difficult to accept (see Waller 2011). Finally there is the group of moral skeptics and “agnostics”, who do not consider the question to be relevant. (*FR*; see also the Stanford encyclopedia article “Compatibilism” by M. McKenna 2004).

⁴⁹ The optimist account according to Strawson is defined by an objectifying attitude toward action within societies. He continues that they tend to point to the evidence about their function and to the fact that there is no evidence in punishment practices that would imply that the thesis of determinism is false, and that as a result they are content to accept the circumstances. He also sketches out one of the common arguments made by the pessimists, i.e., incompatibilists, against the point above in the following steps: 1. Just punishment and moral condemnation are practiced in any given society; 2. This implies truth of moral guilt. 3. Moral guilt implies moral responsibility. 4. Moral responsibility implies free will. 5. Free will implies the falsity of the thesis of determinism. 6. Thus, the thesis of determinism is false. (*FR*, 2.) When all of these arguments are taken together, the result is uncomfortable, as the pessimist stance depends on the truth of 6.

incompatibilists (of the libertarian variety) thus tend to emphasize the control aspect of the moral agent that, as the originator of the action and that the moral desert of the agent, needs to be more prudentially accounted for. The incompatibilist claims that the agent must have the capacity to choose between alternative actions. This, on the other hand, would mean that determinism could not be true. (McKenna & Russell 2007, 2-3.)

McKenna & Russell claim that a key motivation for the incompatibilists is that compatibilists lose something vital, or “deep” from the concept of agency and responsibility (Ibid. 3). This relates mainly to the demand for a capability of control, meaning that the agent is the one who can select one out of possible courses of action. There is however a danger of simplification here. The stance here is that, while the “deeper” role of agency may be a motivation for an incompatibilist position, this desire for or tendency toward a more complex picture of agency does not necessarily lead to incompatibilism.⁵⁰ There are examples of “deeper” statements that do not fall under incompatibilist accounts. The main example of these is Aristotle. For example, according to Sarah Broadie, it would be mistake to call Aristotle a “libertarian”, but she notes that his philosophy can be described as a “proto-indeterminist” philosophy (Broadie 1991, 158). Susan S. Meyer goes a step further, arguing that Aristotle’s view of responsibility,⁵¹ like Strawson’s, is independent of the factual status of determinism, but is not incompatibilist by nature (Meyer 2011). Aristotle’s thought can thus be considered on compatibilist standards.

Strawson discusses his answer to the optimist vs. pessimist debate in *FR*. The argument between the sides is considered further with the help of additional points of disagreement, concluding in the realization by Strawson that both parties fail to see that the disputants are not exactly talking about the same problem: the optimist stops inquiring at a certain point that logic

⁵⁰ See the work of Bernard Williams on this topic (1985, 1993, 1997) and John Fischer and Mark Ravizza’s book (1998), which are discussed here in ch. 4.3.3. Neither, as Eshleman writes, is the division between the two camps necessarily directly indicative of specific views on the concept of moral responsibility any longer: “This general trend of linking the consequentialist conception of moral responsibility with compatibilism about causal determinism and moral responsibility and the merit-based conception with incompatibilism continued to persist through the first half of the twentieth century” (Eshleman 2014).

⁵¹ Perhaps the most cited work on Aristotle within this thesis is Susan S. Meyer’s *Aristotle on Moral Responsibility* originally from 1993, reprinted in 2011, which is used first in chapter 5.1.2 to pinpoint the responsibility for the character side of Aristotle’s moral responsibility. However, as it is discussed in the criticism part of chapter 5.2, it is possible that Aristotle never meant to include the responsibility for character discussion within his theory of moral responsibility; rather it seems that side of the concept would only actually matter for moral education, not responsibility. Aristotle’s theory of responsibility is the topic of chapter 5, where Meyer’s argument that Aristotle’s conception does not conflict with determinism, although she also writes that his character-based view would not be a part of that, is discussed.

allows about the effect of freedom, while the pessimist tries to convince others that what is beyond this point has significance. (FR, 4.) According to Strawson the whole dilemma is based on a misunderstanding:

Both seek, in different ways, to over-intellectualize the facts. Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external 'rational' justification. Pessimist and optimist alike show themselves, in different ways, unable to accept this. (FR, 25, see also McKenna & Russell 2007, 7-8.)

Strawson labels the traditional competing viewpoints as optimistic and pessimistic depending on their attitudes toward the possibility of moral considerations in cases where both states of being are considered.⁵² Optimists claim that whether determinism affects us or not, it makes no difference to the moral choices we make. Pessimists, as Strawson declares, include hard determinists and libertarians alike. These two contradictory positions share, he argues, the negative expectation of morality, as hard determinists deny the possibility of having an effect on one's action, while the latter group resorts to "obscure and panicky metaphysics", including conditions in which moral condemnation or punishment is inappropriate (FR, 1-4).

The pessimist relates to experiencing a negative emotion or a shock, which comes from an assessment made not about the factuality of determinism by itself, but of the perceived callousness of the attitude that the consequentialist (esp. of the economy of threats variety) harbors toward the issue.⁵³ In response, the pessimist seeks a relevant insight, which is the realization that moral condemnation and moral responsibility as its extension is impossible with a fully objective attitude (without emotion)

⁵² McKenna & Russell's assessment of the motivation of the incompatibilist is that the proponents of the stance find the possibility of determinism "depressing or dispiriting". This motivation is mirrored in Strawson's description of "pessimists", the other group he is arguing against, in an analogous way as "optimists", hold a corresponding attitude toward the thesis of determinism. (McKenna & Russell 2007, 4.)

⁵³ Samuel Scheffler (1992) writes on Strawson's contribution in his article "Responsibility, Reactive Attitudes and Liberalism in Philosophy and Politics" that the consequentialist compatibilist forgoes the reactive attitudes and thus bases the entire concept of moral responsibility on clinical, objective attitudes, which in turn shocks the pessimist, who relies more on praise and blame. (Scheffler 1992, 312.)

about the agent and the acts, but also that the incompatibilist fumbles by placing her bets against determinism.

Strawson's contribution to the problem of determinism is found at the end of *FR* by siding (with conditions) with the compatibilists:

Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e. of all we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice. But we do recover it from the facts as we know them. We do not have to go beyond them. Because the optimist neglects or misconstrues these attitudes, the pessimist rightly claims to find a lacuna in his account. We can fill the lacuna for him. But in return we must demand of the pessimist a surrender of his metaphysics. [...] The optimist's style of over-intellectualizing the facts is that of a characteristically incomplete empiricism, a one-eyed utilitarianism. He seeks to find an adequate basis for certain social practices in calculated consequences, and loses sight (perhaps wishes to lose sight) of the human attitudes of which these practices are, in part, the expression. The pessimist does not lose sight of these attitudes, but is unable to accept the fact that it is just these attitudes themselves which fill the gap in the optimist's account. Because of this, he thinks the gap can be filled only if some general metaphysical proposition is repeatedly verified, verified in all cases where it is appropriate to attribute moral responsibility. [...] If we sufficiently, that is radically, modify the view of the optimist, his view is the right one. [...] When we do remember this, and modify the optimist's position accordingly, we simultaneously correct its conceptual deficiencies and ward off the dangers it seems to entail, without recourse to the obscure and panicky metaphysics of libertarianism. (FR, 24-25; 27, my emphasis).

Strawson's concluding answer is that the compatibilist is right, but his position is unacceptable as such. It has to be appended with Strawson's account of the reactive emotions. (*FR*, 27.) The skeptical conclusion (i.e., the rejection of both optimist and pessimist positions), according to which the contradiction of free will and determinism making moral responsibility existing in any form impossible, is counter-intuitive. Strawson argues that the moral skeptic only focuses on the "one-eyed utilitarianism" of the optimist and the "inanity" of the libertarian position. He argues that blame is not a pure metaphysical entity, but an actual psychological feature of human beings as well, and rejecting these facts based on the faults of the viewpoints above would be wrong.⁵⁴ (See *ibid.* 25.)

⁵⁴ "Even if we had some theoretical reason to abandon or suspend these reactive attitudes it would be psychologically impossible for us to do this. To do this would involve "adopting a thoroughgoing objectivity of attitude to others" which is something we are *incapable* of (*FR*, 27-8/69-70)." (McKenna & Russell 2007, 6, compare also with Prinz 2007.)

The problem of determinism comes to a close in Strawson's context: the problem appears "irrelevant" after the focus of responsibility has been shifted toward the reactive attitudes (ibid. 21). As stated, this particular conclusion by Strawson has provoked unprecedented discussion on the topic and its effect on the philosophy of responsibility.⁵⁵ However, not everyone has been persuaded by his result. Strawson's conclusion concerning the problem of determinism is followed at this point of the present work, at least for the sake of the argument. This means that even with its problems, Strawson's solution to the problem is preferable to the continuation of the deadlock. What follows is a detailed, even if not comprehensive presentation of the key concepts in the theory of responsibility found in *FR*.

⁵⁵ Eshleman writes about the continuing debate that Strawson's argument that responsibility ascriptions are oblivious to independent theoretical considerations has not persuaded all of the compatibilists. New versions of so-called "merit-based compatibilism" have been presented by authors such as Fischer and Ravizza 1998 and McKenna 2012. (Eshleman 2014/[2001].)

3.1.2 REACTIVE SENTIMENTS AND BLAME

One of Strawson's most important commentators, Gary Watson, writes of Strawson's emphasis on the reactive sentiments as follows:

All traditional theories of moral responsibility acknowledge connections between [reactive] attitudes and holding one responsible. What is original to Strawson is the way in which they are linked. Whereas traditional views have taken these attitudes to be secondary to seeing others as responsible, to be practical corollaries or emotional side-effects of some independently comprehensible belief in responsibility, Strawson's radical claim is that these "reactive attitudes" (as he calls them) are constitutive of moral responsibility; to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways under certain conditions. [...] There is no more basic belief which provides the justification or rationale for these reactions. The practice does not rest on a theory at all, but rather on certain needs and aversions that are basic to our conception of being human. The idea that there is or needs to be such an independent basis is where traditional views, in Strawson's opinion, have gone badly astray. (Watson 1987, 256-257)

Strawson's refocus treats the entire dilemma as a "pseudo-problem"⁵⁶. The change of focus at which Strawson aimed his work, was the shift from looking at concepts of "freedom" and "responsibility" as analytically definable wholes or solvable problems, toward evaluation of what is happening when we hold someone responsible. This takes the philosophical inquiry away from methodological tools of conceptual analysis toward more a descriptive account of human psychology.⁵⁷ (McKenna & Russell 2007, 4-5 & n9.) The results of this refocus have probably been Strawson's strongest influence on the whole field of ethics. In this way, his "naturalistic turn" involves not just a methodological turn away from conceptual analysis toward moral psychology but also an added emphasis on the importance of emotion in moral and social life in general. (Ibid. 5.)

Strawson's attempt to redirect the discussion of ethics depends especially on the concept of *reactive attitude*. He distinguishes two types of attitude that people hold toward one another: objective attitudes and participant reactive attitudes. According to Strawson, these are essential in our dealings with other people, and the personal emotions we experience when we confront each other depend greatly on them. (FR, 4-5; 13-15.)

⁵⁶ Following Moritz Schlick's "When Is a Man Responsible?" (Schlick 1966/[1939]).

⁵⁷ McKenna & Russell refer to Strawson's phrase "that complicated web of attitudes and feelings which form an essential part of moral life as we know it", a phrase which highlights the importance of social interaction to responsibility (FR, 24).

He describes the *participant* reactive attitudes as “essentially natural human reactions to the good or ill will or indifference of others towards us” (*FR*, 10-11). These reactions are found in the general attitudes and actions that we may take toward each other. The full list of reactive attitudes include personal reactive attitudes regarding others’ treatment of one (including resentment and gratitude), as well as vicarious analogues of these (indignation and approbation), and finally self-reactive attitudes (guilt, shame, moral self-esteem, feeling obligated) (*FR*, 15-16; see also Watson 1987, 259, n4).

The participant reactive attitudes are separate from *objective attitudes*, which we can take toward actors or objects, which we do not qualify as functioning on the same level as we do in terms of moral agency. Participant reactive attitudes are those that are in use when we deal with others who we consider to be of equal standing in terms of cognitive decision-making and consideration. These attitudes appear naturally in that their origin is internal, but the way that they affect our actions becomes refined as we gain more experience. The participant reactive attitudes thus describe our attitudes in interpersonal relationships, first in how they indicate fully developed agents,⁵⁸ or members of the moral community; second in that they adjust the relationships by means of the moral sentiments, among other things, resulting in responsibility relations between agents. This happens for instance through the moral sentiments. In this way, the participant reactive attitudes as well as their vicarious variations appear as emergent phenomena for the demand for the social interactions that are typical of human societies (see *FR*, 16-17).

Participant reactive attitudes solve the problem of determinism, at least for Strawson. What exactly determinism means is left unanswered by him. He tells the reader that an exact answer to the question cannot be attained. This does not, however, he claims, prevent us from focusing our philosophical efforts on the more relevant discussions about “ordinary interpersonal attitudes”. Achieving a clean solution would thus be counter-productive in terms of philosophically interesting issues of morality. The apparent fact that the reactive attitudes are unconnected to the thesis of determinism is the key to circumventing the impasse of the debate. (*FR*, 14.)

Summing up Strawson’s definitions of the key concepts of *reactive attitudes* and *resentment*, he writes that the reactive attitudes are: “essentially reactions to the quality of others’ wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern.” He then argues that resentment is, in his terms, “a reaction to injury or indifference”. (*FR*, 15.) The examination of the excuses results in the distinctions that Strawson makes into the participant reactive attitudes and the objective attitudes that we assume toward each other. (*FR*, 7-13.) The

⁵⁸ Eshleman considers Strawson’s position further, by offering one condition for qualifying as a person: “namely, only they can be proper recipients of the reactive attitudes” (Eshleman 2014).

group of attitudes that is assumed toward the other signifies whether the agent in question is viewed as a full member of the moral community, or whether she is ruled externally, on a decision based for instance on the excuses discussed above.⁵⁹

In the fifth passage of *FR* Strawson describes in greater detail what amount to the most important moral sentiments in his theory: resentment, indignation⁶⁰ and guilt. All of these sentiments, connected to our moral considerations, can attest to the web of relations that involve human beings with each other and, as Strawson claims, they are not subject to the problem of determinism. Strawson considers guilt especially to belong to the group of “vicarious” forms of the sentiments, these emotions being directed at the agent himself or herself. While the other emotions listed might be more relevant in social contexts, both types are equally important for behavior of the agent. The participant reactive attitudes as well as their vicarious variations appear as emergent phenomena from the demands of social interaction. They are common, according to Strawson, irrespective of culture.⁶¹ (*FR*, 14-17.)

Along with the personal and vicarious reactive attitudes, Strawson discusses their counterparts that are described as “self-reactive attitudes associated with demands on others for others.” Feeling bound or obliged is one of these socially oriented attitudes. Others include “feeling compunction; feeling guilty or remorseful or at least responsible, and the more complicated phenomenon of shame” (*FR*, 16). The feelings related to moral responsibility can thus motivate the agent herself to action. What is interesting regarding the reactive sentiments and their classifications is a certain “class of considerations” which reveals some demands for inter-personal retributive measures to be “mere appearances” which Strawson explains works as

⁵⁹ Treating the reactive attitudes as a group of concepts is not a strategy originating with Strawson. He pays homage to the thinkers belonging to the school of thought called moral sense theories, the classical sentimentalists, whose members included David Hume and Adam Smith. Also of note are the writings on “Resentment” by a fellow moral sense theorist, Bishop Joseph Butler (see Butler 1849).

⁶⁰ Strawson indirectly equates indignation with a more forceful form of moral disapprobation (*FR*, 15). This is somewhat comparable with Hume’s idea of blame (see THN, 467-468).

⁶¹ In the final passage of “Freedom and Resentment” Strawson voices some doubts about this. He writes that discussing other cultures might provide criticism of the concept of responsibility used by analytical, Anglophone ethics. For example: “One factor of comparatively minor importance is an increased historical and anthropological awareness of the great variety of forms which these human attitudes may take at different times and in different cultures.” (*FR*, 26.) According to Russell (2013) this is especially true with Wallace’s reading of Strawson, as his way of limiting the concept of responsibility to consist only of negative reactive sentiments makes a concept distinctly biased in favor of “Western, Christianized culture” (Russell 2013, 191).

something that, when applicable to the situation, *inhibits* resentment.⁶² (FR, 17.)

Despite its current reputation as the definitive text on moral responsibility, the exact phrase of “moral responsibility” is actually mentioned at only a few locations in the article, the passage illustrating the considerations leading to resentment inhibition being one of them. Strawson writes: “We may say, stressing the moral, the generalized aspect of the demand: considerations of this group have no tendency to make us see the agent as other than a morally responsible agent; they simply make us see the injury as one for which he was not morally responsible.” (FR, 17.) Here Strawson mentions moral responsibility; the next time “moral responsibility” is mentioned is in the final passage, numbered six.

The conditions requiring resentment inhibition involve situations in which the agent is experiencing hallucinatory delusions, or as Strawson puts it:

But suppose we see the agent in a different light: as one whose picture of the world is an insane delusion; [...] Seeing an agent in such a light as this tends, I said, to inhibit resentment in a wholly different way. It tends to inhibit resentment because it tends to inhibit ordinary interpersonal attitudes in general, and the kind of demand and expectation which those attitudes involve; and tends to promote instead the purely objective view of the agent as one posing problems simply of intellectual understanding, management, treatment, and control. [...] We may say: to the extent to which the agent is seen in this light, he is not seen as one on whom demands and expectations lie in that particular way in which we think of them as lying when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community. (FR, 17-18.)

Strawson’s comment contains an explanation of how the agent can be granted an excuse from responsibility because of psychological abnormalities.⁶³ This kind of suspension from regular inter-personal attitudes is possible according to Strawson even without any particular reasons, including cases in which the reactive attitudes are distinctly moral by nature. As for the situations in which these moral reactive attitudes are completely suspended, the number of motives is limited.⁶⁴ Strawson claims

⁶² This refers to the origin of the idea of the excuses and exemption in Strawson’s context, that some trait or traits of the agent, or a detail or the circumstance can render the blame directed at an offender invalid. Compare further with Wallace’s “blameworthiness inhibitors”, which are discussed below in chapter 4.1.

⁶³ The classifications for these cases are the topic of the next chapter (3.1.2).

⁶⁴ Strawson refers to the considerations that the later authors such as Watson and Wallace call *exemptions*. See ch. 4.1.2 & 4.3.4.

that these total suspensions from the moral reactive attitudes are motivated by weaker motives than those in which the excuses do not lead to a complete shift of perspective toward the other person. This happens in the former cases, in that the agent is no longer treated as a “member of the moral community”, but with objective attitudes instead. While the strain of adversities toward the relationship might be so strong that it would compel a shift of attitude, the benefits of viewing the other with participant attitudes are evident, and this is something that is not nor should be given up easily. Strawson leaves this issue open, claiming that for his purposes the question is not that important; that describing the specific factors affecting the changes of attitude would not need further effort. (FR, 18-19.)

Instead, Strawson reinforces his argument that the reactive emotions of resentment, indignation and guilt are constitutive of moral responsibility:

The concepts we are concerned with are those of responsibility and guilt, qualified as ‘moral’, on the one hand—together with that of membership of a moral community; of demand, indignation, disapprobation and condemnation, qualified as ‘moral’, on the other hand—together with that of punishment. Indignation, disapprobation, like resentment, tend to inhibit or at least to limit our goodwill towards the object of these attitudes, tend to promote an at least partial and temporary withdrawal of goodwill; they do so in proportion as they are strong; and their strength is in general proportioned to what is felt to be the magnitude of the injury and to the degree to which the agent’s will is identified with or indifferent to it. (These, of course, are not contingent connections.) But these attitudes of disapprobation and indignation are precisely the correlates of the moral demand in the case where the demand is felt to be disregarded. The making of the demand is the proneness to such attitudes. Holding them does not, as the holding of objective attitude does, involve as a part of itself viewing their object other than as a member of the moral community. (FR, 23.)

According to Strawson, even the moral skeptic should take note of his position. The usual position of the skeptic,⁶⁵ from where (for instance) the claim “blame is metaphysical” is advanced, is opposed to the optimists’ account.⁶⁶ This is argued against by the former as inadequate, or as leaving something essential about responsibility out. The skeptic is also against the

⁶⁵ The group of alternatives to the skeptical conclusion include: a) that blame is a cognitive judgment, b) an affective notion, which is advocated here by Strawson and c) that it is a conative “operation”. (See Coates & Tognazzini 2014 & ch. 4.2.)

⁶⁶ Eshleman calls optimists referred to by Strawson consequentialist compatibilists (see Eshleman 2014/[2001]).

pessimist.⁶⁷ The incompatibilists are labeled by the skeptic as merely “inane”. Strawson additionally notes the tendency of the moral skeptic to emphasize the inherent confusion of “moral responsibility, guilt and blame”. (FR, 25.)

Strawson’s pessimist says:

that the man who is the subject of justified punishment, blame or moral condemnation must really deserve it; and then add, perhaps, that, in the case at least where he is blamed for a positive act rather than an omission, the condition of his really deserving blame is something that goes beyond the negative freedoms that the optimist concedes. It is, say, a genuinely free identification of the will with the act. And this is the condition that is incompatible with the truth of determinism. (FR, 3. My emphasis.)

The refocusing of the debate from the retributive approaches of responsibility toward the theme of *holding* responsible happens with Strawson pointing out a distinction between “practices or attitudes” regarding punishment and moral condemnation from the agents and the acts that they undertake. While Strawson discusses retribution such that the practices related to it permit (but do not imply) detachment from the agents or actions, which in other words are their objects, he is more interested in “non-detached attitudes”, or the reactive attitudes that people have toward one another. These include attitudes and reactions of people as members of a society, as well as “such things as gratitude, resentment, forgiveness, love and hurt feelings.” (FR, 4-5.) The concepts of attitudes, emotions and sentiments become the essential vocabulary of the further attempts to tackle moral responsibility.⁶⁸ Using these, Strawson gathers the confidence to attempt to “fill in the lacuna which the pessimist finds in the optimist’s account of the concept of moral responsibility” (FR, 22). Strawson’s answer is that the optimist’s account of moral responsibility – appended to his conceptualization of the reactive sentiments – is the way out of the paradoxical situation, in which the philosophy of his time was stuck (see Watson 1987).

Strawson makes note about two possible sources of criticism to his understanding of reactive sentiments, the first of which is related to other cultures and the second to disruptive, new knowledge introduced by empirical research: 1. Strawson writes that increased awareness of historical and anthropological nature may pose problems for his proposed scheme. The

⁶⁷ Eshleman calls pessimists referred to by Strawson merit-based incompatibilists (see Eshleman 2014/[2001]).

⁶⁸ Strawson calls for a return to the ideas upheld by the moral sense theorists: “It is a pity that talk of the moral sentiments has fallen out of favour. The phrase would be quite a good name for that network of human attitudes in acknowledging the character and place of which we find, I suggest, the only possibility of reconciling these disputants to each other and the facts.” (FR, 25-26). This is widely read as a direct reference to the philosophy of David Hume and other philosophers belonging to the group of moral sense theorists, or the classical sentimentalists.

obvious example is the shame-cultures, which instead of guilt and blame emphasize the emotion which Strawson only remarked on as the “more complicated phenomenon of shame” (*FR*, 16). Otherwise, as here, the historical perspective provides an abundance of interesting comparisons (See *FR*, 26).⁶⁹ 2. The other “threat” which Strawson foresees, depending on one’s stance, is the implications that the more empirical sciences can pose for philosophy. While Strawson values the theoretical approach, he expresses the need to address the growing understanding of the human mind that the natural sciences offer.⁷⁰ As for science making the philosophical theories of responsibility redundant, Strawson considers that it is a possibility that can certainly happen, even to the extent that it is a goal for some philosophers to make the related questions disappear, so to speak. (See *ibid.*)

⁶⁹ The Aristotelian theory is further compared below with the Strawsonian example. Notably, Marion Smiley does so with her distinction between classical and modern concepts of responsibility (see ch. 4.3.2).

⁷⁰ Regarding this second type of possible criticism, in the second part of this dissertation the recent advances of the psychological sciences and their effect on moral philosophy are discussed, although briefly, related to the criticism of John Doris (2002) as examined in ch. 6.1.

3.1.3 SPECIAL CONSIDERATIONS OR PLEAS

In the fourth part of *FR* Strawson discusses the extraordinary conditions that affect some of the cases requiring moral assessment. The discussion begins with a reference to “occasions for resentment”, which might mean anything from being harmed intentionally to being hurt by another by accident:

Let us consider, then, occasions for resentment: situations in which one person is offended or injured by the action of another and in which — in the absence of special considerations — the offended person might naturally or normally be expected to feel resentment. Then let us consider what sorts of special considerations might be expected to modify or mollify this feeling or remove it altogether. (FR, 7. My emphasis.)

A significant feature of the reactive attitudes here is their property of being modified or even completely negated by certain circumstances. In other words, these special considerations that modify, mollify, or even cancel out reactive sentiments are studied here.⁷¹ There are two ways that this mitigation might happen: the first is that the person may be excused in terms of her behavior; for instance, if it is determined that the act committed was unintentional, i.e., it was coerced or uninformed. In the other case, the participant attitude may be given up entirely by it being switched to the objective standpoint. In the latter case, the individual in question is no longer viewed as a fully capable member of the moral community and thus not morally responsible.

As we saw in the quote starting this chapter, the occasions when resentment is felt can be affected by special considerations that can “modify or mollify the feeling or remove it altogether” (*FR*, 7). The first group of these that Strawson considers are those that can be discussed with expressions such as “He didn’t mean to”, “He hadn’t realized”, “He didn’t know”, as well as “He couldn’t help it” enforced by claims such as “He was pushed”, “He had to do it”, “It was the only way”, or “They left him no alternative”.⁷² These expressions have in common the feature that they do not ask us to question the “membership of the moral community” or, in other words, the moral status of an adult person that these agents ordinarily have. Rather, they suggest the particular injury as something that is not part of the normal behavior of that agent, but they do not question the agent’s capacity for moral responsibility. (*FR*, 7-8.)

The second group contains two sub-groups, the first being acts that are affected by somehow altered circumstances which diminished the agent’s

⁷¹ The words *excuse* and *exemption* are used in connection with these forms of partial and “full pardon” respectively below following Watson (1987, 259-261) and Wallace (1994, 118, chs 5 & 6). These concepts are returned to in chapter 6.2., where a more detailed exposition is undertaken.

⁷² The acts related to these type-1 pleas are referred here as “coerced or unintended acts”.

capacity to act “normally”. The situations in this subgroup are denoted by descriptions such as “He wasn’t himself”, “He has been under very great strain recently” and “He was acting under post-hypnotic suggestion”. According to Strawson, the first subgroup is of less philosophical interest than the second is. (*FR*, 8.)

The second sub-group is (in this case) more interesting, as it is varied by statements such as “He’s only a child”, “He’s a hopeless schizophrenic”, “His mind has been systematically perverted” and “That’s purely compulsive behavior on his part”. These situations are also ones in which the moral agency or status of the agent comes into question. In these exceptional cases, the circumstances do not deviate from the norm, but the psychology or the moral developmental state of the agent does. Thus, what separates the second sub-group from the first is that the source of this doubt comes not from the environment, but the actor himself or herself. Belonging to this class of actions, our reactive attitudes are a drastically altered. It’s noteworthy that these considerations come into question when the agent belongs to one of the following groups: children,⁷³ the childlike, or the mentally ill: this group in Strawson’s words “allows that the circumstances were normal, but presents the agent as psychologically abnormal – or as morally undeveloped. The agent was himself; but he is warped or deranged, neurotic or just a child. When we see someone in such a light as this, all our reactive attitudes tend to be profoundly modified.” (*FR*, 9.) Strawson uses the example of someone stepping on someone else’s hand: “If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first” (*FR*, 6). In the situation where someone treads on his hand,⁷⁴ Strawson recounts, how we react to the act carries a lot of meaning: if the act was done on purpose we are likely to get angrier than if it was an accident. Strawson notes forgiveness briefly⁷⁵ as a subject, this has received less attention than the more negative reactive emotions of resentment, indignation and guilt. (See *FR*, 6.)

Throughout this work, there are two examples that highlight the phenomenon of blame mitigation and are thus examined from the viewpoint of each author. These are children and the psychopath. Not all writers not take a stand on both of them, but as they do so, it will be noted. Emphasizing his conclusion, Strawson gives an example concerning upbringing: children are increasingly more capable in the sense of being full members of the moral

⁷³ It is perhaps unusual that children and people with psychological abnormalities are discussed side-by-side. Strawson however remains cautious in that he keeps the two groups separated using disjunctions.

⁷⁴ This example is also adopted by Wallace, whose version is looked at in chapter 6.2.

⁷⁵ Garrath Williams has returned to the subject of forgiveness in responsibility in his latest article *Sharing Responsibility and Holding Responsible* (G. Williams 2013).

community, with age, but until they are that, they cannot be fully considered as such. This leads to a suspension of the reactive attitudes as was described above. Examples of children are thus of interest. Eshleman writes on the situation in which the child is initially exempted from reactive attitudes and thus does not qualify for responsibility assessments, but increasingly becomes so through development with age (Eshleman 2014). As another example, Williams also notes on Strawson that “our relations with children occupy a crucial in-between place” (Williams 2010, 8). This example of an ambiguous area of morality poses difficulties for any theory of responsibility.

Strawson says of children:

Thus parents and others concerned with the care and upbringing of young children cannot have to their charges either kinds of attitude in a pure or unqualified form. They are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet truly capable of either. The treatment of such creatures must therefore represent a kind of compromise, constantly shifting in one direction, between objectivity of attitude and developed human attitudes. [...] The punishment of a child is both like and unlike the punishment of an adult. (FR, 20).

Additionally some problems arise with the Strawsonian default model of responsibility with those agents who recognize the criterion of “qualifying as a responsible person” even when they could in fact have a radically different set of morals.⁷⁶ This refers to the case of the psychopaths. (G. Williams 2010.) Strawson doesn’t devote much thought to the examples. He does however touch upon the subject of someone whose “picture of the world is an insane delusion, (FR, 17-18).

⁷⁶ This example is discussed below in ch. 6.3.

3.1.4 THE WEB OF HUMAN ATTITUDES AND FEELINGS

Strawson's attempt to re-adjust the discussion on freedom and responsibility can also be read as an attempt to bring moral philosophy back closer to reality, away from "armchair" metaphysics.⁷⁷ Concluding his call for improved emphasis on emotions as crucial to responsibility, he presents a succinct evaluation of the philosophical practice of his time:

*The object of these commonplaces is to try to keep before our minds something it is easy to forget when we are engaged in philosophy, especially in our cool, contemporary style, viz. what it is actually like to be involved in ordinary inter-personal relationships, ranging from the most intimate to the most casual.*⁷⁸ (FR, 7.)

Adoption of reactive attitudes as the effective mediator of responsibility relations compels a change in the way the relations themselves are perceived. Suddenly, instead of visualizing abstract projections about some transcendent qualities, the sentimentalist inclinations of Strawson's thought suggest another course of action. The evaluations of people about each other carry more weight than, say, thinking about which moral obligations are at work at any particular moment. Instead of asking whether the agent has some property that would positively indicate his or her responsibility, Strawson's way of thinking looks at whether that agent is "a member of the moral community" (see FR, 18). Concluding his elaboration of the ways that reactive attitudes constitute the connections which are relevant to the social appearances of responsibility, Strawson insists that even if we try to ignore our natural reactive attitudes or their import, we would not be able to. (See FR, 19-20.)

Contrasting the objecting attitudes that the traditional compatibilist would have used, Strawson asks us to consider "the web of attitudes and feelings" (FR, 24). The attitudes that we adopt toward each other determines our assessment of their moral status: participant, if the evaluated person is a member of the moral community, objective if not (FR, 9-13). Being a participant in the moral community implies the capability for moral responsibility, for blameworthiness in Strawson's context.

⁷⁷ Strawson seems to be a realist, at least if Ted Honderich is to be believed: "Strawson's common-sense realism, is taken by him to be a real realism, something of which scientific realism is a distortion." (<http://www.ucl.ac.uk/~uctytho/THRIPStrawsonIntro.html>, 20.8.2014)

⁷⁸ It is of course another matter whether Strawson is considered to have succeeded in this. What is thought here is that Strawson was on the right track in terms of refocusing the debate toward more naturalistic terms, but now that the naturalistic turn in ethics he inspire has taken place, he himself appears a little too "cool" and detached for today's standards, but this of course is entirely dependent on one's metaphilosophical inclinations.

Gary Watson has been known since “Responsibility and the Limits of Evil” for coining the term *moral address*, which defined Strawsonian reactive attitudes as a concept that points out an address for blame. He advances similar ideas in his contribution to Coates & Tognazzini (2013) “Standing in Judgment”, but this time from a society’s point of view. His sketch of the idea of “judgmental nonacceptance” describes the social functions of blame and its consequent punishment. He writes about how blaming precedes punishment in a legal context and how it signifies “turning one’s back” on a person who as a result of that punishment is in some way being ostracized from the lives of the others (Watson 2013, 295). Watson’s examples go into very specific details illustrating the trend in how the traditional theories of responsibility (including Strawson’s) have been perhaps too limited in their scope to satisfy philosophers. Thus such details are interesting and more likely than not they will influence the further developments of our theories of responsibility. Whether Strawson was reaching out toward similar points as Watson was is difficult to say. The debates on these contexts have received increasing attention, and they are interesting to follow.

An intriguing subject in the article is the discussion about the *role of society* as it relates to Strawsonian conception of responsibility. There are differing opinions about whether Strawson’s theory is really embedded in actual social interactions, or whether the theory is insufficiently designed to address these settings. According to Gary Watson, the sociality of Strawson’s theory is embedded in his conception of the “participant attitudes”, in that the participant perspective is “inescapable”, since “we [...] could not be led to abandon that framework, *whether or not it is correct*” (Watson 2014, 21, 25). Strawson rekindled the sentimentalist perspective in moral philosophy, emphasizing judgments based on conduct, and even states that responsibility qualifies as moral on the basis of its being a signifier of “membership of a moral community”,⁷⁹ yet the theory leaves the feeling that the agent is observed somewhat in a vacuum. It is argued here that this unconnected feeling comes from the brevity of the paper. According to Watson, who is advancing a similar point: “Strawson misconstrues the space that the [the reactive attitudes] occupy” (ibid. 27).

The sentiment brought forward by McKenna and Russell is also seconded here (see introduction to 3.1). The considerations concerning the social context also lead to a sustained unease about the conception of moral responsibility based solely on *FR*. A way to explain these trepidations is through Strawson’s attempt to bring back responsibility to the social context and away from the theoretical, metaphysical contentions. This creates the sense that he embeds the responsibility evaluations within the “web of human attitudes and feelings” as constitutive of responsibility instead of, say, there being a giant, external, metaphysical red light, rapidly blinking whenever an action committed amounts to wrong-doing. In this sense he

⁷⁹ Strawson 2008, 23. The other property is the retributive role of punishments.

succeeds, but as noted the context is limited. For example, the mechanisms of determining who gets to be considered as part of the “web” (as a node in the web?) is left out of the account. Granted, Strawson briefly acknowledges the difficult cases requiring excuses, children and the mentally ill, and he leaves an opening for criticism in terms of different cultures having an effect the stubborn grip on the perspective of a single subject (or object in case of the excuses) is also notable in his theory. This aspect leads to questions of what emotions affect the agent? What responses face him or her?⁸⁰ And from where do these responses originate?⁸¹ These questions prevent the full picture from forming about the complex society around the agent. Not minding the wild speculation of these concerns, the Strawsonian conception is further bolstered below by the more extensive (but, according to Russell (2013) in some ways narrower⁸²) theory by R. Jay Wallace as well as of some critics.

⁸⁰ If looked at in terms of Bernard Williams’s elements, as is done in more detail in chapter 4.3.4 and in the analysis of chapter 7, all four of his elements of responsibility are firmly anchored to the agent in Strawson’s theory, in which the agent’s causal action is considered (wrong-doing), as well as her intentionality and state of mind (whether she qualifies to be excused), as well as the response (resentment, indignation or guilt, but also blame). The last of these is the crucial part: whether the agent is praised or blamed is considered only from the perspective of the agent in Strawson.

⁸¹ A good summary of the currently popular topic of standing to blame is provided in Macalester Bell’s article “Standing to Blame: A Critique”, in which he writes that “a would-be blamer must have *standing* to blame” (Bell 2013, 263.) This adds a distinctively social dimension to the requirements for responsibility as well. For example, Bell uses the term “contemporary condition”, which requires that the blamer be a part of the same moral community as the agent doing wrong (Bell 2013, 271). Bell highlights a feature that at least Sher and Scanlon, (discussed in more detail in 4.2) have neglected: the standing to blame, that blame is positional and even hostile. As a result, it might be difficult to blame others where one’s own standing is inappropriate for the task. What the origin of the response is is much less discussed. For more information on the recent “standing to blame” discussion, see articles by Cohen (2006), Macalester Bell (2013), Watson (2013) as well as G. Williams (2013). An incompatibilist example on the subject has recently been published by Patrick Todd (2012), in which the topic is approached from the perspective of an omnipotent being.

Additionally Garrath Williams connects “Freedom and Responsibility” to the current discussion about the standing to responsibility or *reproach*, by writing that the questions of responsibility brought about by personal relationships was “a crucial plank of Peter Strawson’s seminal paper, ‘Freedom and Resentment.’ However, while Strawson distinguishes the resentment that an injured party might feel from the indignation that another might feel on the injured party’s behalf,[n4] he does not explore further questions of standing to respond or reproach. For the most part, the literature on responsibility also sets aside this question. Perhaps the most sustained discussion to raise questions of standing is T. M. Scanlon’s.” (G. Williams 2013, 352.) Scanlon’s more socially connected take on blame is discussed briefly in chapter 4.2. “Contemporary debate on blame”.

⁸² Paul Russell’s article “Responsibility, Naturalism and “The Morality System”” in *Oxford Studies in Agency and Responsibility* Vol. 1 (2013) focuses on criticism of the Kantian-influenced “Morality system” aspect found in Wallace’s influential *Responsibility and the Moral Sentiments*. G. E. M.

3.2 SUMMARY

A brief introduction to the theories of responsibility, was presented at the beginning of this chapter, with the cautionary note that in these matters strict scientific precision and conceptual accuracy would be impossible to attain. However, relevant theories of responsibility were identified and Strawsonian and Aristotelian theories on the subject were selected for comparative conceptual analysis. Again, it was noted that the type of comparative analysis here has been done in rather imprecise manner, at least compared to the analyses conducted by the logical positivists and philosophers of language of the early 20th century.

As the most popular theory of moral responsibility is ostensibly based on Peter F. Strawson's "Freedom and Resentment" (2008/[1962]), this article was selected as the starting-point, with which Aristotelian theory as well as the criticisms of the former would be compared. After the brief introduction to Strawson's work, the article itself was discussed, focusing on reactive sentiments and blame, types of excuses and the social context of the theory. Reactive attitudes and more specifically reactive sentiments were the underlying web of social interaction that constitutes moral responsibility in Strawson's theory. Compared to previous writings on moral responsibility, his perspective differed from the previous attempts to define the concept, which commonly claimed that responsibility required a separate, factual judgment independent of the acts being evaluated. Blame as a concept is not explicitly defined in Strawson's article although it is constantly referred to as part of the important reactive attitudes relevant to moral responsibility.

As a particular case demanding special interest, Strawson devotes a passage to the instances of responsibility ascription, which would under normal circumstances lead to blaming the agent, but a crucial difference in agency or circumstance meant that these initial reactive attitudes would require re-evaluation. These situations requiring excusing include unintentional and coerced acts, as well as cases in which the circumstances or the psychology of the agent are altered. These cases will be subject to further examination in the systematic section of chapter 6.

Anscombe and Bernard Williams originally claimed that philosophical writings relying on the sort of morality systems that Kantianism relies on cannot be sufficiently distinguished from the theological roots of such systems. Therefore, something essential is lacking from the theories making claims based on such systems based on secular principles. Wallace is aware of this, arguing that the concern is uncalled for, that using Kantianism as a foundation is possible in his case. The construction of his argument involves a strict limitation of the conception of responsibility in a way that rejects a separate realm of historical, responsibility-related facts. His narrow interpretation of reactive sentiments proves especially problematic according to Russell Wallace's limitations (see Russell 2013, 190).

A question on the social context of Strawson's theory was posed based on the insufficiently defined subjects of moral address and standing to blame or reproach. More recent research on the subject of responsibility has raised the issues about the standing of the blamer, separate from the status of moral agency. The explanation of what the relation between the blamer and the agent is, as well as who has the right, or the standing to blame are not yet directly addressed in *FR*.

4 RESPONSES TO STRAWSON

The fourth section continues the elaboration of the Strawsonian position.⁸³ While the previous chapter was concerned with Strawson's own text, the following chapters are devoted to the current field of philosophy occupied in interpreting and critiquing Strawson's work and its derivations. First R. Jay Wallace's book *Responsibility and the Moral Sentiments* (Wallace 1994) is considered for its careful elaboration of the notions originally appearing in Strawson's "Freedom and Resentment" (2008/[1962]). We should be clear that Wallace's interpretation of some of the central concepts and theoretical structures of Strawson are his own, and in some cases are fiercely debated.

While Wallace's book is currently gaining in popularity as the most extensive rendition of the Strawsonian position, it is not a collection of ideas that have been unanimously accepted. On the contrary, the field of philosophy concentrating on a crucial component of moral responsibility – the concept of blame – has received a great deal of attention from top-class philosophers recently. A brief report on this debate appears in chapter 4.2. While it seems that the jury is still out on the details of what blame is exactly (judgment, emotion, practice or more?), the volume of spirited articles on the subject ensure that important contributions are bound to influence the entire question of moral responsibility in philosophy.

After examining Wallace's important book and summarizing current research on blame, a selection of the most important critical writers on responsibility, almost always in a contending relationship with Strawson's work, are then introduced. As a statement, in this chapter there is not any hierarchy of importance between the authors that would correspond with the order of presentation within this chapter. For example, even though Wallace's work offers the most material on Strawsonian responsibility, its details are not meant to carry more argumentative weight than the briefer criticisms discussed in the later chapters (4.3). Indeed, chapter 4.3.3 specifically, describing Bernard Williams elements of responsibility, is very important here, as the analysis in the latter part of this work relies on them as a framework for the theories of responsibility.

⁸³ The use of the descriptor "Strawsonian" here is in accordance with that of Watson (1987): "My interpretation of Strawson's essay will be in many places very conjectural; and I will sometimes signal this fact by speaking of a "Strawsonian" theory." (Watson 1987, 257, n2.)

4.1 R. JAY WALLACE'S *RESPONSIBILITY AND THE MORAL SENTIMENTS*.

The premier contemporary Strawsonian,⁸⁴ R. Jay Wallace, expands Strawson's reactive sentiments-based view of responsibility in his 1994 book *Responsibility and the Moral Sentiments*, which is one of the most widely cited Strawsonian works on responsibility currently available.⁸⁵ This section (4.1) sets out to describe Wallace's version of the Strawsonian reactive sentiments based conception of moral responsibility. Without repeating what was stated in the previous chapter, the most important similarities and differences between these two versions are noted. The three main points are the reactive-sentiments-based conception of responsibility, excuses and exemptions and reflective self-control, the two first being fairly direct extensions of Strawson's thought, while the last deals with Wallace's own addition, which is influenced by Kantianism.

⁸⁴ Wallace has been called Strawson's main modern follower according to Paul Russell (Russell 2013, 185). Garrath Williams also labels his book a "[h]ighly regarded study indebted to Strawson" (2010). Further support for Wallace's status as the premier follower of Strawson; see, for example, Watson (2014), where Watson comments on Wallace's interpretation in concluding a general assessment of Strawson's influence. (Watson 2014, 27-31.)

⁸⁵ For example, Garrath Williams calls *Responsibility and the Moral Sentiments* "the most well-known and systematic development of Strawson", describing it as a "[h]ighly regarded study indebted to Strawson". Wallace asks when is it fair to hold someone responsible and thus to expose them to reactive emotions such as blame or other sanctions. The response is fair, according to Wallace, if the person has actually done wrong (meaning that he is not excused) and if the person has sufficient mental faculties to comprehend moral reasons. Otherwise the person is exempted from responsibility completely. (See G. Williams 2010.) Wallace's book (1994) is examined here as a direct follower of Strawson's theory on moral responsibility. In addition, his explication of excuses and exemptions (which is borrowed from Watson 1987) is of particular interest here as well, as are the concluding remarks on the concept of reflective self-control.

4.1.1 WALLACE'S STRAWSONIAN RESPONSIBILITY

Wallace's formulation of the concept of moral responsibility as well as its closely related concepts (blame, agent, etc.) and selection of the topics follow the examples in *FR*. Defining moral responsibility in terms of reactive emotions is one of the most important ideas carried over from the earlier work.⁸⁶ The three basic reactive attitudes that Wallace incorporates into his theory are resentment, indignation and guilt (Wallace 1994, 33). The reactive emotions and their relation to responsibility are described in more detail in *Responsibility and the Moral Sentiments*, chapter 3.2, which is titled "Responsibility and the Reactive Emotions". The Strawsonian reactive sentiments are tied to expectations, about which Wallace gives the following description:

[I have] argued that the reactive emotions of resentment, guilt, and indignation should be understood in terms of the quasi-evaluative stance of holding people to expectations. I now want to suggest that this stance provides the key to understanding what we are doing when we hold people morally responsible.

Holding someone to an expectation has been characterized in terms of the reactive emotions: to hold someone to an expectation is to be susceptible to the reactive emotions, or to believe that it would be appropriate for one to feel the reactive emotions, in the case that the expectation is violated. [...] Now it would seem that when we hold people morally responsible we are similarly susceptible to the reactive emotions, if those held responsible breach our expectations, and that we believe it would be appropriate for us to feel the reactive emotions in those cases. This suggests that moral responsibility might be analyzed in terms of the quasi-evaluative stance of holding people to expectations. To hold a person responsible, we might suppose, is simply to hold the person to expectations in the way that is connected with the reactive emotions. (Wallace 1994, 62.)

While other interpretations of Strawson's expectations might have followed a more cognitive or a conative route, Wallace explicitly sticks with the sentimentalist interpretation of reactions to the violations of expectations.⁸⁷ This means the reactive sentiment is at least initially conceived as pure emotion, without cognitive judgment. How Wallace interprets Strawson's

⁸⁶ On reactive sentiments (and a description of the Strawsonian position) see Wallace 1994, ch. 2. "one characteristic that is essential to the reactive emotions, and that may be taken to distinguish them as a class from other types of attitudes [...] is the connection I have described in terms of the quasi-evaluative stance of holding someone to an expectation." (Wallace 1994, 25).

⁸⁷ For comparison, see Coates & Tognazzini (2014), where the authors explicate cognitive, affective and conative varieties of blame. The same tripartite distinction can be utilized with the broader category of reactive sentiments.

guilt in turn is that it is almost invariably directed at the agent himself by himself.⁸⁸ “This [experiencing irrational guilt] suggests the following revision of the initial account: to hold someone morally responsible is to hold the person to moral expectations that one accepts.” (Wallace 1994, 63.)

Even as the distinction between holding responsible and being responsible is better realized in the recent literature (see A. Smith 2007), Strawson can be read as the initial account of moving the debate on responsibility toward the former. However, as is the case with many other distinctions discussed in the article, he does not explain the distinction between holding and being responsible very clearly. In response, Wallace sets out to find, more clearly than Strawson, what it is to *hold* someone responsible (Wallace 1994, 1-2). Furthermore, he wants to keep the question more at the normative level, in practice by inquiring about when it is *fair* to hold someone responsible.⁸⁹ (See Wallace 1994, 5.)

Even as Wallace is skeptical about the prospects for an understanding on a definite view, labelled “the Strawsonian account of responsibility”, he takes up the challenge of spelling said view out. In writing about Strawson’s direct influence on him, Wallace explains that the sentimentalist nature of his idea of responsibility is more important than Strawson’s theory:

Thus there is no fixed and stable view that might be labeled the Strawsonian account of responsibility. Strawson’s original lecture contains a wealth of ideas, and the many philosophers who have been influenced by the lecture have naturally chosen to develop and defend different ones among them, and to develop them in different ways. For my part, what I find promising in Strawson’s position are not so much the official answers he provides to the question of the compatibility of responsibility with determinism, as the suggestions he makes about the connection between responsibility and the moral sentiments. (Wallace 1994, 10.)

Chapter 3 of *Responsibility and the Moral Sentiments*, titled “Responsibility” begins with a description of the situation of the conversations on freedom and responsibility in philosophy. According to Wallace, holding people responsible has not been adequately explored in the history of the field. The main feature of accounts of moral responsibility includes the understanding that the concept primarily refers to the ideas of blame and moral sanction and the practice of holding one another morally responsible. In other words, the idea of holding morally responsible is often interpreted as preparedness to blame or sanction based on moral offenses and the attempts to limit them.

⁸⁸ Strawson implied this kind of self-blame by using the phrase “vicarious” alternatives to the reactive attitudes (see Strawson 2008, 15-16).

⁸⁹ Thanks to Teemu Toppinen for kindly pointing out the difference between *holding responsible* and the idea of the *fairness of holding responsible*, the latter of which Wallace is trying to define in his book (1994).

As specific as Wallace's description is, and despite leaving something important out, this account seems to be a popular way to define moral responsibility,⁹⁰ but he notes that "philosophers have not yet given us a satisfactory interpretation of [the subject of holding one responsible]." (Wallace 1994, 51.)

The core of Wallace's theory is the claim that:

To hold someone morally responsible is to view the person as the potential target of a special kind of moral appraisal. People who are morally responsible are not seen merely as acting in ways that happen to be good or bad; they are not just causally responsible for certain welcome or unwelcome happenings [...] Rather, the actions of morally responsible people are thought to reflect specially on them as agents, opening them to a kind of moral appraisal that does more than record a causal connection between them and the consequences of their actions. (Wallace 1994, 52.)

In general, Wallace is careful in keeping his account firmly in the realm of the normative (i.e., "the normative interpretation"). His use of holding responsible exemplifies tactics used in the book to avoid the confusion that Fischer and Ravizza (1993, 18) and Angela M. Smith (2005) mention. Another level of detail in Wallace (1994) is what Paul Russell describes as following: "Wallace is well aware that his narrow construal of the reactive attitudes commits him to an interpretation of moral responsibility understood entirely in terms of the conceptual resources of "the morality system"⁹¹ (Russell 2013, 189, referring to Wallace 1994, 39-40 & 64-66.) These delineations fix Wallace's focus on causal responsibility, such as responsibility for action and accountability approaches to the subject; however, the moral sentiments and the special nature of interaction in their terms are kept in mind throughout the book. (See Wallace 1994, 53.)

Equating moral responsibility with moral sanctions can lead to behavioristic positions. According to Wallace, moral responsibility tied to moral emotions improves the concept from the bare bones of the "economy of threats" approach. As he writes: "[the reactive sentiments account] treats the stance of holding people responsible essentially in terms of attitudinal conditions, and so avoids the behavioristic danger of associating responsibility too exclusively with moral sanctions." (Wallace 1994, 66-67.) The way that the economy of threats approach views responsibility leaves a lot to be desired, a point on which all the later compatibilists agree. To this, Wallace adds that this view could also explain the hard determinist account

⁹⁰ Compare this statement, for example, with T. H. Irwin (1980), which includes a rendition of Aristotle's moral responsibility formulated in similar terms.

⁹¹ "The Morality System" in Russell's article refers to Bernard Williams' criticism of the (Western, Christian in origin) idea of any kind of system of external morality in his book *Ethics and the Limits of Philosophy*. (See Bernard Williams 1985.)

of why people should not be retributively blamed or punished, because they do not have a choice in the matter.⁹²

Instead Wallace proposes the schema: “(N) S is morally responsible (for action x) if and only if it would be appropriate to hold s morally responsible (for action x)” (Wallace 1994, 91). Following up, Wallace elaborates on the requirements of holding responsible:

First, we need to know what it is to hold someone morally responsible, what is involved in taking up this stance. Second, we need to specify the norms by reference to which the appropriateness of that stance is to be gauged. We already have an answer to the first of these questions in hand, as the reactive account [...] tells us what it is to take up the stance of holding someone morally responsible. [...] To render the normative interpretation more determinate, it will be necessary to specify particular substantive norms by reference to which the question of the appropriateness of holding people might be answered. (Wallace 1994, 92.)

In his 2014 article published in *Oxford Studies in Agency and Responsibility* vol. 2, titled “Emotions and Relationships: On a Theme from Strawson”, Wallace continues to build on the Strawsonian theme, discussing among other subjects the concepts of resentment and injury in further detail. Wallace separates three aspects of resentment: injury, quality of will and demand. (Wallace 2014, 126.) Injury can be interpreted according to Wallace in two ways: 1. injury can be taken to mean simply harm. 2. injury can be constituted, and Wallace thinks that the second interpretation is closer to Strawson’s original meaning. (Wallace 2014, 128-129.) Injury, as a cause of resentment and harm is a result of action and the avoidance of it can be rationalized as a function of responsibility. How an agent is evaluated as responsible is a big question that Wallace works on in *Responsibility and the Moral Sentiments*. One particular, yet important result of this evaluation is excusing. How the excuses are described by Wallace is explained in the next section.

⁹² See also appendix 2 “Alternate Possibilities”, where Wallace argues that incompatibilists usually present an argument with two stages: First, that responsibility depends on the possibility of having alternate possible actions to choose from. Second: that as this is not the case, when determinism is true, responsibility is invalid.

4.1.2 EXCUSES AND EXEMPTIONS

In talking about excusing, Wallace refers specifically to excusing acts that would evoke moral responses in an ordinary situation. At the most immediate emotional level, this means excusing from *blame*. As quoted before, Wallace describes the act of holding one responsible as amounting to “view[ing] the person as the potential target of a special kind of moral appraisal” (Wallace 1994, 52). This entails more than just a direct, causal responsibility for acts that “happen to be good or bad” (ibid.). Actions of these fully responsible agents display a property of that agent that allows for the sort of evaluation that moral responsibility requires. In other words, responsibility involves a *disposition* belonging to the agent, which is necessary for moral sanctioning. (See Ibid. 56.)

Following H. L. A. Hart,⁹³ Wallace seems to think that “People who are morally responsible may be made to answer for their actions, in the sense that their actions render them liable to certain kind of distinctively moral responses” (ibid. 54; see also 124). What this means is that acts committed by the agent are required for the responses such as praise or blame, or of those signifying other reactive attitudes.⁹⁴ Regarding the responses, where Strawson was reticent about blame, Wallace ascribes more importance to the word. He holds blame to be the “most salient” of the moral responses, as they indicate situations in which the agent has actually done wrong. Only the agents capable of a morally responsible status are viable recipients of the appraisal of this sort. (Wallace 1994, 54.)

The implications of applying Wallace’s schema to action provoke thought. What are emphasized are *excuses*, referring specifically to Austin’s distinction between justifications and excuses.⁹⁵ Wallace states that: “[E]xcuses are unlike justifications in blocking or inhibiting responsibility for whatever action was performed.” (Wallace 1994, 120.) While justification can apply to both right or wrong action, excuses seem to apply solely to the cases of presumed wrong-doing. Excuses, then, are a counter to this initial judgment or reaction, as Wallace explains: “Excuses, by contrast, aim

⁹³ Herbert Lionel Adolphus Hart has worked mainly in the 60’s on the retributive side of the concept of moral responsibility. Hart offered his own distinction of moral responsibility, which discussed the concepts of liability, accountability and the effect of roles in society. (See esp. Hart 1968.) According to Garrath Williams, Hart’s “rule-utilitarian account of excusing conditions in the law has been influential to contractualist accounts of responsibility, especially T. M. Scanlon’s.” (G. Williams 2010, 11). It is a regret that Hart’s writings are unable to be referred to any further extent here because of their different focus.

⁹⁴ This response can then work to drive the actual sanction through in the social context. At this stage, however, the theory is not concerned about the succeeding events.

⁹⁵ Austin’s remarks on the excuses, which are found in his article “A Plea for Excuses” (Austin 1956/57), are returned to briefly in chapter 6.2.

precisely to challenge the claim (or suspicion) that s was morally responsible for x" (Wallace 1994, 121).

In order to "work up Austin's hypothesis [against the incompatibilists] into an account of excuses, therefore, we need an explanation of the normative force of the excuses, as he characterizes them." (Wallace 1994, 122.) The excuse in Wallace's case acts as a "*blameworthiness inhibitor*".⁹⁶ Some account of why the initial moral response, consisting of blame, needs to be re-evaluated and principles of resulting annulling of moral sanctions is required. Following Austin, "there is a connection between doing x intentionally and being morally responsible." So "doing x intentionally is ordinarily a condition for responsibility for x, such that if s has not done x intentionally, then it is unfair to hold s morally responsible for x." Wallace asks "Why would it be unfair to hold s responsible for x if s has not intentionally done x in the first place?"⁹⁷ (Ibid.)

Wallace describes further the notion of blameworthiness inhibitors by referring to the influence of Austin and Strawson in that "excuses serve to show that an agent has not really done anything wrong" (Wallace 1994, 127). Wallace argues that excuses show "[i]n Strawson's terms, [...] that s's action, though it may have appeared to violate our moral demands, did not really violate any moral demand that we accept." (Ibid.) Thus Wallace describes excuses as a concept that distinguishes the causal responsibility from blameworthiness by accepting the former and calling the latter into question. Wallace expands Strawson's two categories into four distinct groups of excuses, to which chapter 5 of the book is devoted. These categories, which are listed as the headings of the sections of the chapter, include the following: "Inadvertence, mistake or accident"; "Unintended bodily movement"; "Physical constraint" and "Coercion, necessity and duress". He adapts the "someone treading on my hand" example from Strawson in giving a description of each. (Ibid. 136-147.) "Inadvertence, mistake or an accident" are descriptive of situations in which the agent is uninformed, at least in advance of the consequences of his or her action. Wallace writes briefly notes that in such cases the "blameworthiness will be highly localized". As an example, Wallace gives the case in which there is a fire next door, but the agent doesn't know it, so he omits the act of extinguishing the flames (ibid, 137). "Unintended bodily movement" is a pretty self-explanatory category. Nevertheless, Wallace describes the group of excuses with some examples and a statement that "in these cases it seems quite clear that there is no intentional action at all" (Ibid. 140). "Physical constraint" refers to those cases in which one is unable to control or restrain oneself from acting because of force applied to the agent from an external or internal origin. In

⁹⁶ Wallace's *blameworthiness inhibitors* resembles the idea of *resentment inhibition*, presented by Strawson. (See ch. 3.1.2.)

⁹⁷ In this context, Wallace indicates Hume's conception in THN 2.ii, containing a sketch of responsibility, which accepts character based concerns.

the case of physical constraint, writes Wallace, that omission to act cannot be considered a violation of one's moral obligations. (Ibid. 142-143.) "Coercion, necessity and duress" are a group of excusing conditions in which harm is caused by the agent because of avoidance of another (imposed) instance of harm;⁹⁸ the example Wallace gives is a bank teller who gives the money to the robber to avoid being shot (ibid. 143). In concluding, Wallace writes the "whenever people are genuinely excused from responsibility for their actions, those actions will not have been morally wrong" (ibid. 147). Strawson described excuses in terms of special conditions or pleas that could have the effect of "modifying or mollifying" the reactive sentiments, but he also mentioned pleas that could, in circumstances which were certain, eliminate the reactive sentiments entirely. To help distinguish this group of excuses with maximal effect, Wallace adopts the distinction of excuses and exemptions, the latter referring to these strong cases of adjustment of the reactive sentiments, from Watson's article "Responsibility and the Limits of Evil". Wallace describes the key difference between the excuses and the exemptions as follows:

Exemptions are like excuses – and unlike justifications – in granting that doing x intentionally is not morally impermissible. By contrast to excuses, however, exemptions do not inhibit responsibility for a particular action x by providing grounds for doubting that s did x intentionally. [... W]hereas excuses inhibit responsibility for a particular act by showing that a morally accountable agent has not done anything morally impermissible in the first place, exemptions block responsibility for a particular act by showing that an impermissible act has been done by someone who is not, in general, a morally accountable agent. (Wallace 1994, 156.)

A particular feature that carries over from *FR* to *Responsibility and the Moral Sentiments* is that the exemptions are discussed as specific groups that (by belonging to that group) are exempted. While Strawson was careful to keep the groups of children and the other agents who were subject to the "special considerations" separate, Wallace puts children in with five other groups of exempted agents. (Ibid. 166-180.) Wallace describes the groups receiving exemptions in chapter 6: children (ibid. 166-167), the insane or mentally ill (ibid. 167-171), addicts (ibid. 171-175), the hypnotized or otherwise controlled (ibid. 175-176), psychopaths (ibid. 177-178) and the extremely stressed (ibid. 179-180).⁹⁹

⁹⁸ The topic has been recently demonstrated quite clearly (albeit in a consequentialist environment) in an article titled "Moral Coercion" by Saba Bazargan (Bazargan 2014).

⁹⁹ Wallace's groups are fairly straightforward derivations for Strawson's examples in *FR*. Watson (1987, 260) describes Strawson's examples as: "being psychotic, being a child, being under great strain, being hypnotized, being a sociopath ("moral idiot"), and being "unfortunate in formative circumstances".

Without giving an exact description of what “insane or mentally ill” means, Wallace refers to the main quality of these conditions which he considers the cause of the exemption. He argues that the members of this group lack the means for reflective self-control (ibid, 167). What reflective self-control means is looked at in the next section. Addicts include all types of people with uncontrollable desires, but especially “willing” addicts (ibid, 171). The fourth exempted group is a bit surprising, albeit with a precedent in Strawson: the hypnotized or otherwise mind-controlled. Wallace states that these cases involve desires “implanted” by other agents, meaning that the desires which are not one’s own can lead to the intentionality of the act being questioned. However, as Wallace admits, implanted desires cannot be grounds by themselves for exemption, as people are constantly “put under the spell” of advertising. This group of exemptions is also subject to the loss of reflective self-control. (See ibid. 175.) On psychopaths, Wallace begins by describing problems that are not necessarily the results of psychopathy. Nevertheless, Wallace suggests, again, that psychopaths could be exempted because of disabled reflective self-control (Ibid. 177). Finally, Wallace writes that the extremely stressed might alternate between the excuses and exemptions, but what decides the outcome is the effect of reflective self-control (ibid. 179). What is this reflective self-control? This final point of interest in Wallace’s theory will be discussed in the next section.

4.1.3 REFLECTIVE SELF-CONTROL AND THE MORALITY SYSTEM

In Wallace's book (1994) *reflective self-control* is something that agents belonging to the exempted groups lack. Presented as his own version of the Strawsonian solution to the problem of determinism, acts without excuses that are accompanied by reflective self-control are thought by Wallace to be immune to concerns raised by determinism. He sees reflective self-control as "the rational power to grasp moral reasons and to control one's behavior by the light of them" (Wallace 1994, 2). It's a strong definition of the concept therefore, close to moral agency.

Wallace's account of reflective self-control, which is based on a synthesis of Kantian and compatibilist philosophy, is fairly general overall and not specific.. He writes that: "powers of reflective self-control are to be construed as forms of broadly psychological competence or capacity, like the general ability to speak a language, or to read musical notation and reproduce the music read on an instrument" (ibid.186). According to Wallace, in terms of moral responsibility these general powers (which are unfortunately not detailed at much greater length) are required from agents¹⁰⁰ in order for them to be able to meet moral obligations; however, the opportunity to exercise these obligations is not needed (ibid.186-187). He describes the concept through the example of young children, who by definition lack reflective self-control, as they have not been able to develop it yet:

Young children to begin with, are clearly not taken to have developed fully the powers of reflective self-control. Cognitively, they are at the stage where they are learning how to apply moral principles and the often very complicated concepts (such as harm or reciprocity) that the principles incorporate. Affectively, they have not yet acquired the ability to control their behavior reliably in accordance with such moral principles. [...] Reflective self-governance requires the capacity to comply with moral obligations, and young children may not yet have the ability to act from motives of this sort. In that event their

¹⁰⁰ Wallace takes precautions to secure his position somewhat by leaving the exact substance of "general powers of self-control" undefined (see Wallace 1994, 192). One example of what these powers might actually consist of is given from the perspective of psychology by Kotabe and Hofmann, who write in their article "On Integrating the Components of Self-Control" (2015) that: "self control can be understood as an evolutionary adaptive function that has emerged from the need to coordinate motivations resulting from multiple dedicated subsystems in the brain." The authors focus on *desire*, *higher order goal*, *desire-goal conflict*, *control motivation*, *control capacity*, *control effort* and *enactment constraints* as seven distinct components of self-control. (Kotabe & Hofmann 2015, 621; Kotabe and Hofmann refer to Livnat and Pippenger (2006).) Thanks to the "Self-control and Moral Enhancement" workshop organized by Susanne Uusitalo and Polaris Koi in Turku for this reference. Most of these components have analogues in philosophical theories of responsibility as well (compare with Aristotle's *boulesis*, *telos*, *prohairesis*, *dunamis*, etc.) and it would be a very interesting project to compare these historical contexts with each other.

compliance with moral obligations will be due to quite different sorts of motives: the wish to avoid punishment, say, or the desire for parental approval. [...] Because children lack these abilities, or are still in the process of acquiring them, it would be unreasonable to hold them fully accountable with respect to the moral obligations that we accept. (Wallace 1994, 166-167.)

The conception clearly occupies a central position in Wallace's theory, as all of his examples of exempting relate to the theme of reflective self-control (ibid.). Are all Wallace's exemptions based on lack of reflective self-control? Wallace explicitly claims that each of these groups is exempted because conditions related to an insufficient or diminished amount of reflective self-control (see ibid. 167-179). In addition to children, exempted groups include agents such as the mentally ill who, according to Wallace, suffer from conditions affecting reflective self-control (ibid. 167-170). Addicts, who consist of all types of people with uncontrollable desires, are treated in a similar way throughout his examination: while alluding to the debate on whether there is a difference between willing and unwilling addicts, he "brackets" the issue and notes relatively simply that addicts lack the capacity for reflective self-control (ibid. 170-171). The hypnotized have this property manipulated for them, and even the problematic state that moral philosophy has trouble getting grips with, which the psychopath inhabits, is explained by Wallace as having a "disabled" faculty of reflective self-control (ibid. 175-177). The fourth exempted group, while a bit surprising, has a precedent in Strawson (1962): the hypnotized or otherwise mind-controlled.¹⁰¹ Wallace states that these cases involve desires "implanted" by other agents, meaning that desires not being one's own can lead to the intentionality of the act being questioned. However, as Wallace admits, implanted desires themselves cannot be grounds for exempting, as people are constantly "put under the spell" of advertising. The fourth group of exemptions is also subject to the loss of reflective self-control. (See ibid. 175.) On psychopaths, Wallace begins by describing some problems that do not necessarily represent the result from psychopathy. In this case Wallace suggests, again, that psychopaths could be exempted as a result of disabled reflective self-control (ibid. 177). Finally, Wallace writes that the extremely stressed might alternate between the excuses and exemptions, but what decides the outcome is the effect of reflective self-control (ibid. 179).

Wallace seems to imply that this distinctive quality related to these conditions is the cause of the exemption. For example, without saying exactly what "insane or mentally ill" means, he claims that the members of this

¹⁰¹ Wallace's groups are fairly straightforward derivations for Strawson's examples, which are given in *FR*. Strawson's examples are described by Watson as: "being psychotic, being a child, being under great strain, being hypnotized, being a sociopath ("moral idiot"), and being "unfortunate in formative circumstances". (*FR*; Watson 1987, 260.)

group lack the means for reflective self-control (ibid, 167). The main problem with his way of presenting the groups of exempted agents is that it is simply a limit imposed on the application of the theory of moral responsibility in the form of a list. It can be argued that this practice of shutting off the entire assessment of responsibility is too strict when it is apparent that the agent is not capable of self-regulating to a full adult degree.¹⁰² This instrument of exclusion might have to be appended.

The importance of reflective self-control is further emphasized by Wallace in instances such as him writing in response to Fischer & Ravizza (1993) that:

On my view, by contrast, responsibility for an action or omission requires that the action or omission reflect some quality of choice [...] But so long as an agent retains the general powers of reflective self-control, it is not required that the choice for which the agent is held responsible is actually the result of deliberation on the basis of reasons at all. [...] It would be unreasonable to hold people to obligations that are supported by reasons in this way [which promotes compliance] unless they possess the general powers of reflective self-control: the power to grasp and apply moral reasons, and to control their behavior in accordance with them. (Wallace 1994, 190-192.)

Instead of relying on classic compatibilist conditional analysis, according to which the agent must have chosen in order to be held morally responsible, Wallace considers “the general powers of reflective self-control, and the absence of specific excusing conditions” as the decisive conditions of responsibility (Ibid. 194.) The requirement for reflective self-control indicates that none of its contradictory exemptions exist. Instead of choices and the result of the conditional analysis, Wallace states that:

What makes it fair to hold s responsible are the different kinds of facts that I have tried to characterize [...] about the content of the choices expressed in s’s general powers of reflective self-control” (ibid.) and again: “the fundamental condition of accountable moral responsibility is not freedom of the will, but possession of the powers of reflective self-control.” (Ibid. 230).

The selection of these exempted cases seems just to cater to the need to highlight the desired properties of reflective self-control. The explanation for exempting is limited in the case of mental illnesses to declaring the lack of that feature, by which in relating the concept to “defective reason” Wallace avoids details on the subject with a disclaimer about “complexities [being] beyond the scope of [his] discussion”. (Wallace 1994, 167.)

¹⁰² Critics such as Kelly (2013) take this viewpoint. Her suggestion of adopting a more gradual understanding of the excuses is discussed at the end of chapter 6.2.

Criticism of Wallace

The details of Wallace's theory are many, and are not examined here in any further detail. Suffice it to state that Wallace's responsibility consists of (a) the subject having the powers of reflective self-control as well as (b) a lack of excuses (see Wallace 1994, 154-155). Whether excuses in this case include the exemptions is unclear. Since Wallace seems to indicate that the exemptions are more related to (a), the highlighted groups of exempted agents categorically lack reflective self-control in his context (see *ibid.* 166-180.)

By emphasizing the realization that reflective self-control is a significant contributor to the cause for most exemptions, Wallace implies that the lack of reflective self-control could be the causal reason for every exemption. On the one hand, Wallace writes: "Possession of these powers [of reflective self-control] is thus a basic condition of the fairness of holding people accountable, [...] I try to show that all the standard exemptions can be construed as impairing or depriving a person of these conditions of accountability." (*Ibid.* 155.) Wallace starts out with reflective self-control as his goal, arguing that reflective self-control is a key component of accountability, with the counter-examples of the exempted groups enforcing his point.

Later on, diverging from his initial position, Wallace appeals to the "conventional understanding", as he also writes: "the most important exempting conditions as they are conventionally understood, seem to impair the powers of reflective self-control." (*Ibid.* 180.) Thus, having presented the exempted groups, he seems to be advocating historical facts according to which the exempting conditions that define the members of the groups of exempted agents and said membership are causing the lack of reflective self-control.

By emphasizing the realization that reflective self-control is a significant contributor causing most exemptions, Wallace seems at least to imply that reflective self-control could be the causal reason, or at the very least a necessary requirement, for every exemption discussed. If this implication is valid, it is suggested here that Wallace is *affirming the consequent*.¹⁰³ The

¹⁰³ *Affirming the consequent*: $a \rightarrow b, b \rightarrow a$

En (sx): s is exempted (based on n) from a response to having done x.

R(s): s has powers of reflective self-control.

E1 (sx): s is exempted based on being a child from a response to having done x.

E2 (sx): s is exempted based on suffering from mental illness from a response to having done x.

etc.

$\exists s \forall x (En(sx) \rightarrow \neg R(s), \neg R(s) \rightarrow En(sx))$

lack of reflective self-control might be a sufficient condition for exemption, but it is questionable whether lack of self-control is a necessary requirement for it. Lack of self-control is a feature related to these groups, but it does not precede them. The exempted groups themselves just happen to have a family resemblance by which lack of self-control can appear related to them independently, but where reasons and type of feature vary. Thus, in the case where exemptions explain lack of reflective self-control, implying obstacles to holding an agent morally responsible, the original context in which the exemptions were described as reasons for mitigated reactive sentiments, can be read as fallacious in the sense above.

Wallace remains careful regarding the topic of historical responsibility (discussed more below in chapter 4.3.4). He labels the concerns of historical responsibility under the “conditions of autonomy”, which he claims “should not be confused with moral responsibility” (Wallace 1994, 53). In other words, in Wallace’s view the concept of moral responsibility does not include the sort of historical concern that Aristotle, for example, deals with in terms of character in the *Nicomachean Ethics* III.5 (see chapter 5.1.2). This non-historicism seems to relate to the argument based on choices made in the past by the incompatibilists.

Wallace also writes on the effect of previous events:

I cannot see how to make sense of the idea of a prior and independent realm of moral responsibility facts [...] the practice of holding people responsible is characterized by a highly structured set of emotions and actions, namely the reactive emotions and the blaming and sanctioning behavior that expresses them. But it seems incredible to suppose that there is a prior and independent realm of facts about responsibility to which such emotions and actions should have to answer. (Wallace 1994, 88.)

Instead of assuming a “prior and independent realm of facts” Wallace opts for an internalist position, according to which moral facts are considered dispositional, or as parts of their relevant actions and events. The dispositional stance on morality leads to an emphasis on a “current-time-slice”¹⁰⁴ understanding of responsibility, meaning that prior events leading to the assessment of moral facts are dismissed as non-existent. In this way, Wallace’s more in-depth sketch of the Strawsonian conception of responsibility also sharpens the shortcomings of the position. Wallace alleviates these problems by allowing for “facts about responsibility” to manifest internally, just without the separate layer of a “prior and independent realm” (ibid. 92). Compared to Strawson’s more

¹⁰⁴ Fischer and Ravizza (1998) utilize the phrase “current-time-slice notions”, which is originally derived from Michael Bratman’s and Robert Nozick’s work (Fischer and Ravizza 1998, 178-181, see also chapter 4.3.3).

intersubjectively defined “web of human attitudes”, Wallace’s dispositional internalism seems even more atomist by nature.

Further detail to Wallace’s position appears in his more recent publications. In *Responsibility and the Moral Sentiments*, Wallace touched on the notion of the internalism vs. externalism of morality, which he continues in his 2013 article “Rightness and Responsibility”. In the former, Wallace notes that Strawson’s attempts to reconcile the pessimists and optimists rely on external argumentation (Wallace 1994, 101). In the article, Wallace elaborates the positions and defends internalist accounts of morality. According to the internalists “there is a nonaccidental connection between [moral rightness and motivation].” The internalists claim that a judgment about rightness is required to motivate a person to act the rightly, while the externalists deny this. Wallace advances the claim that the internalist position is also relevant in the case of responsibility in that “rightness and responsibility are noncontingently connected” (Wallace 2013, 224).

Others have noted Wallace’s uneasy mix of Strawsonian and Kantian influences. A notion that received only (3) passing mentions by Strawson and that Wallace devotes more time to flesh out is *moral obligation*, including the sibling concepts of moral right and moral wrong. This notion related to responsibility is of course criticized by authors such as Elizabeth Anscombe and Bernard Williams as being dependent on “the morality system” (Wallace 1994, 64). Wallace argues against Williams’s claim, that moral obligations would be based on “a distinctively theological conception of ethics”. Instead, according to Wallace the moral emotions are independent of such obligations, or more accurately, free of the need of being held to such obligations (ibid. 64-66; see also B. Williams 1985).

Paul Russell’s description in his critical 2013 article “Responsibility, Naturalism, and “The Morality System”” published in *Oxford Studies in Agency and Responsibility Vol.1* touches on two of these points:

Wallace’s compatibilist account weaves together two distinct strands of thought. The first is a broadly Strawsonian description of holding people responsible, interpreted in terms of our reactive attitudes (Wallace 1994: 8-12). The other strand is his Kantian theory of reflective self-control or moral agency (1994:12-15.) [...] Taken together, these two strands constitute what Wallace calls his “normative interpretation” of responsibility, which maintains that the correct way to understand what it is to be a morally responsible agent is by way of describing those conditions under which it is fair to hold an agent responsible (1994: 5, 15,64). [...] While these moves are generally consistent with Strawson’s original approach, Wallace aims to substantially modify and amend this approach by providing

*a narrower and more fine-grained account of moral reactive attitudes.*¹⁰⁵ (P. Russell 2013, 187-188.)

The term “morality system” is derived from Bernard Williams’s *Ethics and the Limits of Philosophy* (see Williams 1985). In chapters two and three, Wallace describes the concept of morality system, generally positively. He discusses less how this compares with the “prior and independent realm of moral responsibility facts” that he rejects (see Wallace 1994, 88). Generally his view is in accordance with the original Strawsonian position that moral responsibility is not dependent on an objective judgment as a requirement for the moral responsibility assessments. It could be argued that within the idea of the morality system there is room for a prior, but connected realm of moral responsibility facts (i.e., Watsonian historical responsibility) in his theory.¹⁰⁶

These last two points about reflective self-control and the morality system can be looked at as Wallace’s own expansion of his replicated, but expanded Strawsonian position to that point. The Kantian influences in particular have provoked criticism. For example, Russell (2013) writes that Wallace takes a narrower path with the reactive sentiments than Strawson. This leads Wallace’s Strawsonian/ Kantian position insufficiently¹⁰⁷ sensitive to cultural variations in reactive sentiments (see Russell 2013, 197-199).

Michael McKenna (1996) zeroes in on the contrast between Wallace’s efforts to put the Strawsonian schema clearly on the normative level and his

¹⁰⁵ Russell’s article is determined to show why this “narrowing” of the reactive attitudes divorces Wallace’s version of the concept from the reality. Russell’s argument is supported here in terms of the outcome; his subject of criticism in itself is not commented on here, but other factors indicate the same conclusion.

¹⁰⁶ In his article “Responsibility and the Limits of Evil” (1987) Gary Watson advances a number of very good questions that criticize the Strawsonian conception of responsibility. He points out the lack of historical or biographical information related to the reactive attitudes as assessments of responsibility. Such demands on moral responsibility are usually delivered from the direction of the incompatibilists, but as Watson opens up the question, he brings up Aristotle as well as leading to a sort of denial of Strawson’s claim that moral responsibility would not be affected by determinism. There is no resolution to the problem within his article, but the issues he raises are still relevant (Watson 1987, 274-280). Another good example demonstrates that responsibility and blame are not always intertwined: “if one shares a moral fault with another, one may feel it inappropriate to blame the other. Here the point is not that the other is not responsible or blameworthy, but that it is not *one’s* business to blame.” (Watson 1987, 283). Watson’s article is discussed in section 4.3.4.

¹⁰⁷ Russell points out that “Given that [Wallace’s] narrow construal insists that moral responsibility be understood in terms of the concepts of the morality system, and that this interpretation alone constitutes genuine or real moral responsibility, it is compelled to exclude all other understandings (i.e., as based on a broader construal of ethical reactive attitudes) as falling outside the parameters of moral responsibility.” (Ibid. 199.)

more general tendency toward moral agency (in the naturalist context) not having any metaphysical extensions/implications:

Perhaps the most interesting and controversial aspect of Wallace's book is the link he builds between the normative interpretation of the free will debate and his rejection of the metaphysical picture of moral agency. Wallace rejects a conception of the debate as an attempt to ascertain the metaphysical facts about moral agency that obtain "prior to and independent of our practice of holding people morally responsible" (87). This complements his normative interpretation of the debate, in which he accounts for moral agency by appeal to the conditions that make it fair to hold an agent morally responsible. Some might remain uneasy about giving up the plausible picture provided by the metaphysical view. On this view, the practice of holding morally responsible is to be tempered by the conceptually prior question of what makes a person a moral agent. One wonders whether Wallace's normative interpretation of the debate depends upon his rejection of the metaphysical picture. If it does, Wallace's position might seem to some more controversial, but yet less plausible. (McKenna, 1996.)

The contradiction described by McKenna is most apparent where Wallace denies the existence of any separate realm of moral facts in explaining how facts can be conceived either as metaphysically prior entities or as fixed dispositions (Wallace 1994, 88-9). Furthermore, Scanlon and Nagel have criticized Wallace's theory of moral responsibility as being only agency- and character-based, as opposed to the more complex theories that the philosophers have advanced for their own part, which are relevant and interesting in their own right, but are not looked at here.

What Wallace's refinement of the Strawsonian position does is that the previous authors explication of the various situations leading to the excuses, and the various agents qualifying for the exempt positions clarifies the concept of the excuse, about which Strawson initiated discussion (by following Austin). *Responsibility and the Moral Sentiments* is certainly an ambitious and arguably hitherto unsurpassed formalization of Strawson's theory. As observed, the viewpoint is not without its problems. As Russell criticizes Wallace, the theory propounded in the book contains a few drastic delineations that alter the course set by Strawson's original paper (Russell 2013, 187-188 & 203-204).

The limitations of being concerned only with normal adults, generic mental states, and giving up on the praise-side of responsibility, not to mention the contradictory claims about the metaphysical layers of the theory of responsibility, present their own challenges. Compared to Strawson and the relatively brief sketch of moral responsibility in *FR*, Wallace's theory in *Responsibility and the Moral Sentiments* has attracted much more systematic effort to articulate the sentimentalist, normative account. The cost

of this refinement however has been an even narrower group of applicable moral agents, and a distinct reliance on the functions of reflective self-control as well as the Kantian notion of the morality system. All in all, while the themes discussed in this chapter are the substance of Wallace's contribution, they are difficult to assess thoroughly here. The main reason for this is the admitted omission related to Kantianism (discussed in chapter 2.) meaning that these features fall outside of the scope of this effort.

Resolving to read Wallace (1994) as a realization of the Strawsonian position while leaving out the Kantian-influenced advances, which are very impressive in their own right, is a choice taken here. Especially important as a topic that is much more fleshed out in Wallace (1994) than it is in *FR* is the nature of the concept of blame and its interplay with the questions of responsibility in general. Perhaps through Wallace's input, the theme of blame has now been "promoted" to be the main topic of the philosophy of responsibility. This topic, *blame*, is the subject of the next section.

4.2 DISCUSSION: THE CURRENT DEBATE ON BLAME

Blame has been a popular topic in Anglophone moral philosophy for quite a while, and the revival of naturalistic philosophy, led by Strawson's example, has inspired some of the most prominent philosophers, such as T. M. Scanlon and George Sher, to take part in the discussion. As has been mentioned previously, Coates & Tognazzini present an essential selection of articles on the topic in *Blame: It's Nature and Norms*. (2013).¹⁰⁸ Emphasis on these authors is reinforced by Coleen Macnamara's 2013 article in the same book. According to her, there are "multiple accounts of blame", the currently influential ones including:

1. Sher's blame: "blame is a set of behavioral and attitudinal dispositions that have their source in a belief-desire pair: the belief that someone has acted badly and the desire that the one blamed not have done what she did or not have the character traits she has."
2. Scanlon's blame: "to blame is to change one's comportment toward another in acknowledgement of the fact that she has done something to impair your relationship with her."
3. Strawsonian blame (incl. Wallace, Wolf and Darwall): "I blame another when I respond to her conduct with resentment, indignation or disapprobation, that is I blame another when I respond to her wrong-doing with a negative reactive emotion."
(Macnamara 2013, 141.)

Further refinement is found in the *Stanford Encyclopedia of Philosophy* article "Blame", also by Coates and Tognazzini (2014). For example, the authors distinguish the conception of blame as discussed in the context of theories of moral responsibility from "causal or explanatory blame", which is not interesting as a philosophical question.¹⁰⁹ Both of these sources (2013 & 2014) give an overview of the currently on-going discussion about blame. In

¹⁰⁸ Coates & Tognazzini's 2013 compilation *Blame* is a collection of high-quality articles, written by the most important thinkers on the subject. It includes texts by Wallace, Scanlon and Sher, to name a few. Darwall's work is in the curious position of being usually grouped together with the Strawsonian approach (see Macnamara 2013, 141), but in terms of blame mitigation he advocates the use of character. George Sher, T. S. Scanlon and Angela Smith, also listed in the latter group, are referenced below. Based on this book, the editors have also produced the *Stanford Encyclopedia of Philosophy* article "Blame" (2014).

¹⁰⁹ The example used to describe causal blame goes as follows: "It is this notion of blame that is at stake when we say that Hurricane Hugo is to blame for the destruction of Charleston's harbor, or that the cat is to blame for knocking over the vase." (Coates & Tognazzini 2014.)

the book (Coates & Tognazzini 2013), the authors go through various interpretations of what the concept is actually about and where it matters. They have noted that, indeed, most researchers working on blame approach the concept from some larger perspective such as theories of moral responsibility, as is the case with Wallace (1994). The article makes more general claims such as that blame is taken to be as “a common feature of our shared moral experience” (Coates & Tognazzini 2014). In both publications by Coates and Tognazzini, a division of the theories of blame into three categories is based on the common three-part separation of the mind in psychology.¹¹⁰ These groups that distinguish the dominant features of each of the theories best are called “cognitive”, “emotional” and “conative” theories of blame. Cognitive theories of blame emphasize blame’s primary role as a judgment or an evaluation. Emotional theories on the other hand see blame as primarily having an affective quality. Finally, the conative group of blame-theories view blame more as a directing force such as desires or expectations that people hold. (Coates & Tognazzini 2014 & 2013, 8-15.)

Strawson, who defines the group of emotional theories of blame, is the starting-point, but alongside him, the important theories of other philosophers are discussed, especially those mentioned above. This aside, it would be a mistake to claim that this is the first time in history that the concept of blame has been discussed within Anglophone analytic philosophy. In fact, as early as 1956 Austin noted (literally, in a footnote) that the topic of blame has been examined thoroughly in a way that does not call for continuity. His exact words referred to blame as a “well-flogged horse” (Austin 1956/57, 7, n2). Of course, Austin’s notice comes from a different age, when the main concern of the professional philosophers was to determine the relations between natural and formal languages, and it is thus taken in that context. In this instance, Austin contrasted the conflating use of blame at once as causal blame and other times as the key component of responsibility, as is discussed here.

Something of the opposite is the case with the concept of praise. According to Wallace especially, praise is not genuinely related to blame, but is rather a disanalogous counterpart used in the same contexts. While traditionally paired with praise (since Aristotle), philosophers who have been previously working on moral responsibility have more recently concentrated on the concept of blame at the expense of the former. Wallace is the first writer examined here who explicitly claims that “praise does not seem to have the central, defining role that blame and moral sanction occupy on the practice of assigning moral responsibility.” (Wallace 1994, 62). This coincides with the general sentiment that the interest of the responsibility theorists has

¹¹⁰ Technically Coates and Tognazzini gather additional theorists under a fourth category of “functional theories”, such as Hieronymi (2001), Matthew Talbert (2012), Victoria McGeer (2013) as well as Michael McKenna (2012, 2013). As these form quite a broad selection of theories from different angles, the study of these contributions is left to another occasion.

increasingly centered on the negative repercussions of wrong-doing, sanctioning and even retributive punishment in some cases.¹¹¹ This is not to say that praise has been satisfactorily explained away. On the contrary, Coates and Tognazzini predict that praise will be the next big topic of philosophical inquiries in the coming years.¹¹²

Today, however, the Strawsonian naturalistic revival and the developments that followed it have taken an interest in blame. The fresh perspectives on the subject, based on the writings of the notable philosophers working particularly in ethics rather than the linguistic analyses of past is pursued here. Some of these recent contributions are discussed, including Wallace, Scanlon¹¹³ and Sher. The contributions of these three seem to be crucial to the current efforts to establish a common understanding about nature of blame.

¹¹¹ See, for example, Gary Watson (2004/[1996], 283), who writes that “We seem to have a richer vocabulary of blame than of praise”.

¹¹² E.g., “[Macnamara’s article (2013)] suggests that a full understanding of praise will require an independent inquiry, which is something very few philosophers have undertaken.” (Coates & Tognazzini 2013, 26.) While Wallace excludes praise, the concept is present in at least the interpretations of Michael S. Pritchard, who emphasizes the ideas of moral development and education in his book *On Becoming Responsible* (1991). Similarly to the way Dennett classifies the “plateau” of moral agency, Pritchard quotes Thomas Reid, who held that there is a certain threshold of development that is reached at some point in life, after which the agent becomes a full-fledged moral agent. Blameworthiness and praiseworthiness are mentioned in this context as being important to the status of moral agency. (See Pritchard 1991, 33.) Another recent attempt at focusing attention on the praiseworthiness-aspect of responsibility is by Andrew Eshleman in his article “Worthy of Praise” (Eshleman 2014b). Praise is also part of Doris’s characterization of moral responsibility, also related to the context of moral education. His version is very critical, despite being about the whole idea of character-based moral education (Doris 2002, 121-127). Nevertheless, he includes consideration of moral education within his sketch of the subject, a context in which praise has an important role. Doris laments the lack of evidence regarding the benefits or efficacy of the character-based educational efforts, but one of the scarce actual findings of empirical research is Grusec and Redler (1980), according to whom attributional praise is more effective than unattributed praise when prosocial behavior is encouraged in children (Doris 2002, 126). Thus, at least in this way both Doris and Pritchard acknowledge praise as an element of accounts of moral responsibility. Granted that as Doris begins his evaluation of the character-based systems of moral education, he explicitly links the topic with a contemporary American “character education movement”, which is not known internationally, but does nevertheless sound unpleasant (see Doris 2002, 121). This movement might provide one reason or motivation for Doris to criticize character in the first place.

Wallace has recently been criticized by Paul Russell (2013) for framing his conception of responsibility too narrowly. This includes the overall emphasis of Wallace’s theory on the negative reactive attitudes.

See also Andrew Eshleman’s “Worthy of Praise” (2014b) as an example of a recent publication on the subject.

¹¹³ Thanks to Simo Kyllönen for drawing attention to Scanlon’s relation to Strawson’s philosophy.

Coates and Tognazzini (2013 & 2014) present cognitive theories of blame as the point of comparison with all other types of theory that are discussed here. These relying on the idea that blame is primarily a judgment. They write of cognitive accounts that they “all capture something deep and important: blaming involves *evaluating*.”^[n22] When we blame others, we see them as having dropped below some standard that we accept (or perhaps that we think they should accept, whether of excellence, morality or respectful relationships.” (Coates & Tognazzini, 2013, 9.)

Some adherents of cognitive theories of blame go as far as to claim that the concept can be reduced solely to its cognitive aspects, as the mid 20th century utilitarian J. J. C. Smart did (1961 & 1963). Blame would then fundamentally be conceived as “a judgment or evaluation that we make about an agent in light of her attitudes or her actions.” Smart, however, represents the “ledger view” type of responsibility theory which, according to Watson and Eshleman (discussed above in chapter 3), forgoes the social side of responsibility including notions of *aretaic appraisal* and *reactive sentiments* as constituents of responsibility. For this reason, his point of view is not considered as interesting in this case. (See Coates & Tognazzini 2014.)

Necessarily, as Coates & Tognazzini state, supporting a cognitive theory of blame does not result in a position that advocates the “ledger view” (ibid.). Another cognitively aligned blame theory, which they offer as an example is that of Pamela Hieronymi (2004), whose approach is appealing in stating that blame is primarily a judgment. The other theory discussed, emphasizing the cognitive aspect, is T. M. Scanlon’s earlier work (1986)¹¹⁴. Both Hieronymi’s article and Scanlon’s earlier theory locate blame within the realm of judgments, and more specifically judgments about the dispositions of others toward us; for example, the judgments that are made of another where that person is acting out malice toward us.

Hieronymi’s cognitive account of blame builds on Gary Watson’s attributability/accountability division.¹¹⁵ She does this by emphasizing the attributability “face of responsibility” (in Watson’s words), according to which our moral assessments of each other, including those leading to blame, involve an evaluation of that person and the action he or she takes based on what we consider important. Watson calls this practice the *aretaic appraisal* (see Watson 2004/[1996], 271), and Hieronymi goes as far as to claim that the reactive attitudes cannot be separated from this notion of *aretaic appraisal* (Hieronymi 2004, 39).

¹¹⁴ It is interesting to note that Scanlon’s early theory differs from his latter work. According to Coates and Tognazzini’s article, in his work Scanlon’s early view of blame (1986) is conceived as primarily being cognitive by nature and that in his later work (2008 & 2013), his views shift toward a conative position.

¹¹⁵ See introduction to chapter 3 in this work.

Some of Gary Watson's criticism of the Strawsonian position is concerned with blame.¹¹⁶ Coates & Tognazzini take his writings to belong within the cognitive theories. They highlight his use of the term "aretaic appraisal"¹¹⁷ which, in addition to the determination of the blameworthiness of the agent, is included in responsibility assessments: not only do people decide whether to blame one another or not, we compare our conduct in comparison with the ideals that we would wish to uphold. Watson is critical of the emotional theories of blame, such as those advanced by Strawson and Wallace, but he does not dismiss the importance of the reactive sentiments. Rather he seeks to expand the quite limited core of the Strawsonian outlook by calling out its deficits, which need appending.

In his well-known article "Responsibility and the Limits of Evil", Watson mentions Gandhi and Martin Luther King as examples of people who act against harmful practices essentially by blaming these practices in order to eliminate them. The example shows that the basic functionality of blame as proposed by Strawson as "a retributive act[,] one that involves a withdrawal of good will", is not the only way to utilize the concept. This demonstrates that blaming is not valuable for its own sake, but just as a negative attitude. (Schoeman (ed.) 1987, 22 quoting Watson 1987, 286.)

If these theories were applied to an actual case, each one would emphasize a different way to react to wrong-doing. Say that A runs over a dog, which is owned by B. The way that B would blame A for the incident would depend on the theory. If cognitive, the act of blame would be interpreted as a judgment, i.e., B would cognitively make the judgment or the evaluation that "A ran over my dog. That's reprehensible. The person, A, is to blame for this."

The Strawsonian theories offer the emotional interpretation instead, according to which blame is primarily an emotional response to wrong-doing. Strawson's additions, which have been discussed above, concern equally blame-focused theories (when compared with theories of responsibility). Coates & Tognazzini explicitly mention Strawson as the key figure of the emotional theories of blame as his "'Freedom and Resentment' is the founding document of contemporary work on blame" (Coates & Tognazzini 2013, 5). In fact, in the introduction to their book, the same group is labeled "The Strawsonian Account" (3, 13), while in Coates & Tognazzini (2014) it is designated "Emotional Theories of Blame". In these theories, blame and the capacity to receive it are the main constitutive elements of responsibility in the Strawsonian sense.

Strawson is equally credited as a reviver of the discussion but, as was the case with responsibility, he did not provide a specific definition of blame in "Freedom and Resentment" (see Coates & Tognazzini 2014). In this work however, the main proponent of Strawsonian blame is considered to be R.

¹¹⁶ Watson's contribution to discussion on responsibility is considered more closely below in chapter 4.3.4.

¹¹⁷ See also Watson 2004/[1996], 266-267 on *aretaic appraisal*.

Jay Wallace, whose theory is focused on blame as an emotional response or a reactive emotion. This is a natural reaction to observed wrong-doing, wherein the acts do not carry a separate metaphysical (binary) condition,¹¹⁸ these responses, as well as the attitudes related to them defining or being interpreted as moral responsibility.

Macnamara's article "Taking Demands out of Blame" describes the Strawsonian group as including Strawson, as well as Wallace, Wolf and Darwall (Macnamara 2013).¹¹⁹ Describing how blame works according to the Strawsonians, she writes: "I blame another when I respond to her conduct with resentment, indignation or disapprobation, that is I blame another when I respond to her wrong-doing with a negative reactive emotion. (Macnamara 2013, 141.)

If based on an emotion-based theory of blame, such as Wallace's for instance, A in the example above would instantly react instinctively, non-cognitively. For example, in a Strawsonian context the observation and the realization of what A's action had brought about would fill B with resentment, making B proverbially "sick to his/her stomach", and would provoke a reactive emotion toward the transgressor A in a forceful act of condemnation.

The conative accounts according to Coates and Tognazzini are those that "emphasize motivational elements, like desires and intentions, as essential to blame." George Sher (2006) and Thomas M. Scanlon (2008, 2013) are mentioned in the article as showing the most developed conative theories of blame. Sher's criticism is aimed at both of the other types of theory, in that he does not believe that emotion is required for blame. Instead, he argues, it should be equally based on a cognitive element as well as a conative element, resulting in an account by which blame is ultimately conceived as a belief-desire pair. Scanlon's theory equally disposes of the emotional, and offers instead a cool, rational account of social relations and appraisals that occur. (Coates & Tognazzini 2014.)

Sher's blame is interesting,¹²⁰ as the concept in his theory necessarily contains both a cognitive and a conative aspect. Blame, in his theory as

¹¹⁸ For the sake of the argument here, it is assumed that Wallace's reliance on the normative level is not considered as such. However the notion is questioned elsewhere in this work (ch. 4.1.3).

¹¹⁹ Victoria McGeer also argues for a psychologically informed account of blame in her article "Civilising Blame" (McGeer 2013).

¹²⁰ The *other* recent contribution to the topic of blame is by George Sher as he discusses in *In Praise of Blame* (2006) and "Wrongdoing and Relationships: The Problem of the Stranger" (2013). His version criticizes Scanlon's blame, proposing instead that blame is actually a belief-desire pair, a concept that brings together cognitive and conative aspects of the phenomenon. However Sher's account is accused by Victoria McGeer (in 2013's "Civilizing Blame") and Christopher Franklin of being too "sanitized" (Coates & Tognazzini 11). Whether or not these two accounts of blame gain a permanent place alongside Strawson's version, these contributions have ensured that the topic is alive and well in the current academic discussions.

described in his book *In Praise of Blame* (2006), is conceived as a belief-desire pair. Blame in his account consists of both an understanding that a wrongdoer has committed an offence and how, as well as an idea of how this person has failed to live up to the moral expectations. This was the cognitive side. The conative side of blame consists of a retrospective desire that the blamer has. This desire is aimed at the wrongdoer such that the blamer wishes that the wrongdoer had decided otherwise, or avoided the act. Based on these two aspects, Sher claims we can better understand the moral relations between us.

One of Sher's theory's proponents is Angela Smith, who describes Sher's blame as follows:

Blame should be understood as a set of dispositions to have certain attitudinal and behavioral reactions, and these dispositions should be understood as traceable to a single desire-belief pair that includes (1) a belief that the person in question has acted badly or has a bad character, and (2) a corresponding desire that the person not have acted badly or not have a bad character (Sher 2006, p.112). (Smith 2013, 33.)

Thomas M. Scanlon's more recent writings have seen him modeling a theory of blame¹²¹ that has had a mixed reception. The intricacies of his theory are impossible to discuss satisfactorily here, but it is possible to briefly demonstrate how his conception of blame is unique: Scanlon's blame is distinctly social; it always involves the act of evaluating fellow human beings and the alteration of expectations toward them based on the former. Thus, if person A offends B with an improper act, B can evaluate what has taken place and she can lower her further expectations toward A as a result. In the context of Scanlon's theory, Coates and Tognazzini maintain that: "to blame someone is to recognize, and make modifications that express that one recognizes, that things cannot go on as before with that person. The relationship has been impaired [as a result of actions done by the other]" (Coates & Tognazzini 2013, 12).

In Scanlon's view in his "Interpreting Blame" (2013) a "person is *blameworthy*, [...] if he does something that indicates intentions or attributes that are faulty by the standards of a relationship". He continues: "A judgment of blameworthiness is a judgment that an action shows something about the agent that impairs, in various ways, his or her relations with others." (Scanlon 2013, 88.)

¹²¹ The Scanlonian interpretation of blame is discussed in his book *Moral Dimensions: Permissibility, Meaning, Blame* (2008) among other works, but here in most detail. In "Interpreting Blame" Scanlon's brand of blame appears as "a class of responses to morally faulty actions" (Scanlon's 2013), which at least according to Susan Wolf is too "mild" an interpretation, especially as contrasted with her "Blame, Italian-style" presented in the article with the same title in the book discussing Scanlon's philosophy, *Reasons and Recognition* (Wolf 2011).

Scanlon's blame is thus distinctly a social concept by nature, involving multiple agents, and based essentially on relations (in an abstract sense) and the dynamic of modification of these relations (whereas Strawson's blame could be abstractly conceived as being projected without necessary repercussions). Scanlon's concept of a relationship is defined as "a set of intentions and expectations about our actions and attitudes toward one another".¹²² (Scanlon 2013, 86.)

While many have praised Scanlon's intricate and fundamentally social account of blame, his conceptualization of the term has left others, especially the advocates of the emotional theories, wanting. For example, Coleen Macnamara has defended Scanlon's view, while R. J. Wallace has dismissed his attempt by stating that Scanlon's "dispassionate opprobrium" is not blame at all (Wallace 2011).

Conative theories need a bit of explication: If Sher's example were to be followed, the event (in the example about the dog being run over) would bring about a strong desire in B toward a resolution that is different from the accident, based on the belief that A should not have driven over his or her dog. Scanlon in turn would say that A through her negligence would not have met B's expectation of not having her dog run over. This would thus provoke B to further alter her expectations of A about A's behavior.

From these examples it is easy to imagine how the broader theory of responsibility could be affected with a "strategic" change in the interpretation of blame. It is, for example, certain that Watson's criticism (looked at in the next chapter), which states that Strawson's conception of responsibility is flawed in that it excludes the most "evil" of human beings by disqualifying them from moral assessment by means of exemptions (Watson 1987). Going with a cognitive theory of blame could make sense in this case in that the individual's wrong-doing is easier to *judge* to be harmful without any mixed emotional content concerning the agent. This would be the case in situations where extraordinary agents, such as people with memory-related illness or psychopaths, etc. are the actors.

The proponents of the cognitive accounts of blame have been critical of the emotional counterparts. For example, Coates & Tognazzini (2014) note that Hieronymi's theory advances the viewpoint that blame is something more important than just the emotional accompaniment of a judgment, and that an affection accompanied by said judgments would be little more than distractions (see Hieronymi 2004, 121). Similarly, as an advocate of a conative theory, Sher goes after the emotional theories: "Sher (2006) argues that emotional responses are unnecessary for blame. For example, Sher argues that we can blame a loved one without feeling negative emotional reactions." (Coates & Tognazzini 2014.) Sher claims that the cognitive

¹²² See also G. Williams 2013, 352: "Scanlon understands blame as essentially a response within a relationship, when that relationship has been impaired."

judgments about doing wrong are accompanied by the conative desire concerning that how the person evaluated would not have acted or decided to act in the way in which he/she did. However, this belief-desire pair can bring about the negative emotions that are usually associated with blame, but it is the belief-desire part that takes precedence. Pamela Hieronymi responds to Sher in her 2008 paper “Sher’s Defense of Blame”, by claiming that Sher’s model, even while worthwhile in its description, gives insufficient evidence that “blame” might be constituted as such a pair. (Coates & Tognazzini 2014.) Scanlon’s theory has recently attracted some of the most polarized comments. Scanlon’s version of blame is too “dispassionate” compared to Strawsonian blame, at least if Wallace and Susan Wolf are to be believed (see Wallace 2011; also Wolf 2011). According to Wallace, the emotional response serves an important feature that is incorporated into blame: judgment changes meaning as a result (Coates & Tognazzini 2014). Sher asks in his article, which criticizes Scanlon’s theory “Wrongdoing and Relationships: The Problem of the Stranger” (2013) how Scanlon’s relational theory of blame can take a stranger into account? If the wrongdoer is not known to the person offended, how can the expectations be altered? This is especially the case if there are no prior expectations without the framework of pre-existing mutual reactive attitudes, or Sher’s own retrospective desire for that matter — how can the Scanlonian blame-as-adjustments-to-relationships function? (see Sher 2013, 49-51.) Sher concludes the article by stating that the Scanlonian account (as well as Linda Radzik’s by extension) is untenable:

[B]ecause the views [on relational approaches to blame] discussed here are among the best worked-out representatives of that approach and because each must either rely on implausible empirical claims or else retreat to a nonempirical and potentially question-begging account of the relevant relationships, I think it is safe to say that the prospects are not too bright. (Sher 2013, 65.)

Some commentators still favor the emotional accounts: Victoria McGeer in her essential article “Civilizing Blame”¹²³ in Coates & Tognazzini (2013) sketches “anger characteristically manifested in blame [is] a feature, moreover, that [...] plays a crucial role in regulating behavior by way of making salient the demands that shared norms place on our actions and attitudes.” (McGeer 2013, 183.) By carefully presenting evidence about the effect of the psychology of people when acting, McGeer argues for the emotional side of blame as the more plausible, but also drafts a way forward by describing a way of accommodating these affective notions of the concept in the context of academic philosophical discussion. This is in accordance with an earlier comment on Scanlon’s theory by Susan Wolf in her article

¹²³ The article additionally contains actually interesting connections between Scanlon, Sher, Wallace, and to the empirically-minded moral psychological writings of Prinz (2011) et al.

“Blame, Italian Style” (Wolf 2011). Both commentators criticize the overly dispassionate Scanlonian take on blame, while making an attempt to emphasize the importance of the emotional response in the practice of blaming.

If there’s one precise trend in the current debate, it is that for many writers the Strawsonian account is unable to satisfactorily explain blame, and alternatives must be sought. Angela Smith presents a defense of Sher’s and Scanlon’s work on the concept of blame in her article “Moral Blame and Moral Protest” (Smith 2013, 47). In her article “Taking Demands out of Blame” Coleen Macnamara (2013) also argues for Scanlon’s and Sher’s approaches, especially where Strawsonian blame is under the magnifying glass of Macnamara’s article. In part, Macnamara writes in response to Wallace that it is not always so that holding someone responsible would be retrospective blame of the other due to a particular misgiving. Rather blame (as well as responsibility as a result) extends to prospective cases as well. She gives the example of a pizza box, in which a demand written on the box works as a prospective form of holding the persons reading the demand responsible. She separates the evaluative and the deontic realms of morality, claiming that Wallace operates in the latter with his idea of holding someone responsible. In her view, it is important to recognize the connection between demands and the negative reactive emotions. (Macnamara 2013, 160-161.) Finally, Gary Watson gives an interesting inversion of terms in Watson (2013), where he writes about non-judgmentalism in his contribution to Coates and Tognazzini (2013), “Standing to Judgment”. This is related to “the standing to blame” discussion initiated in his earlier article (1987), but interesting as it is as a development, this train of thought on non-judgmentalism will not be followed further here.

The current debate on blame is interpreted here as follows: Scanlon’s theory is an admirable attempt to describe responsibility and blame as relationships between people. Indeed, this relational account of blame seems attractive precisely because the *social* is integral to responsibility. However, it is thought that what Scanlon describes as blame, with its reduced emotional content and its emphatic connection with relationships, is deemed problematic. Sher’s example of the stranger displays a similar problem within Scanlon’s framework as is the case with Strawson’s account, which compromises the moral status of less-than-ideal agents completely. For the reasons examined, Scanlon’s account of blame (and responsibility as its extensions) seems too limiting, in that it is very likely able to describe the situations that it is designed to describe without error, but moved to less appropriate circumstances will fail in execution because of unspecified details (as was the case in Sher 2013).

Sher’s thinking is followed here, at least in the sense that responsibility is considered to be multi-faceted. Blame in Sher’s theory consists of the dual qualities of desires and beliefs. This applies to responsibility as well. Although Sher is convincing, it is thought that his account might be

incomplete at present and that it should be developed a little more through the academic discussion.

Strawson's "default view" of blame is considered to be the content of the concept of blame from now on, but just for the sake of the argument. Strawson's account is problematic, and some of its shortcomings are discussed in the following pages. Nevertheless, in the rest of this work blame is still considered to be primarily constituted by reactive emotions.

There's a lot more to the discussion related to blame, (including a set of technical details about the procedures related to blaming in Coates & Tognazzini (2014), and lamentably these are not looked at here in any more detail.¹²⁴ Considering the theories mentioned above, it is too early to draw conclusions within the boundaries of this work about who or what theory is correct about blame. All of them touch on important issues, and it is easy to imagine that if one of them were nailed down as the right one, the larger picture of responsibility would accommodate to it as a result. Each of the participants has brought up interesting points about the complex ground that the concept of blame covers. There might yet be refinements and even syntheses based on work by current theorists, which might provide an even more compelling description of blame. Perhaps a further effort will combine the relative conceptual lightness of the Strawsonian account, the social aspects of Scanlon's work and the accuracy of Sher's theory. In any case, "blame" is a topic that will certainly be interesting to follow for the next couple of years.¹²⁵ For the sake of the argument, however, the Strawsonian account (including Wallace's refinements) is used here as a rule-of-thumb, but the compelling points made by the proponents of the other theories will be contrasted with the status quo when relevant.

In the next section, centering on criticism, Herbert Fingarette will be looked at as a completely different take on the concept of responsibility.¹²⁶ Marion Smiley's book presents a key problem in the current discussion of responsibility, while Bernard Williams's studies on the subject sketch a framework that will be utilized further in this work. Finally, Gary Watson's article in Schoeman 1987 is mentioned as a criticism of the Strawsonian default picture, and his work will be given attention in the last section, where many of his important points are addressed.¹²⁷

¹²⁴ See Coates & Tognazzini 2014 for an introduction to themes such as blameworthiness in relation to blame as well as aforementioned procedures such as "proportionality", "warrant", etc.

¹²⁵ There have already been some developments in this discussion in the published volumes of *Oxford Studies in Agency and Responsibility* (Shoemaker (ed.) 2013, 2014 (with Tognazzini) & 2015).

¹²⁶ Herbert Fingarette's *Mind* article "Responsibility" (1966), which originates from the same period as *FR*, serves here as an important contrast to the latter in that "responsibility" can be understood in many ways.

¹²⁷ Watson helps to distinguish (but does not go deeply into) the key interpretation relevant to this work: he notes chapter III.5 of the *Nicomachean Ethics* and, related to it, discusses *historical responsibility*, which is relevant in the case of the *responsibility for character* topic discussed more in

4.3 CRITICISM

In this section, some of the most important alternative views on the nature of moral responsibility are looked at. First, Michael McKenna & Paul Russell's *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment"* (2007) is discussed as a work that puts *FR* directly under the magnifying glass and offers very poignant criticism. After this, more general questions about moral responsibility are asked by Herbert Fingarette (1966), Marion Smiley (1992), Bernard Williams (1993 & 1997) and finally Gary Watson (1987 & 2004).

The final section of "Critical Themes Concerning 'Freedom & Resentment'" by McKenna & Russell (2007), contains a couple of noteworthy details about the background of and the discussion of the article. It seemed to be the case that at the time (mid-20th century) it was thought within the school of analytical philosophy that the free will problem could be resolved by conducting a thorough conceptual analysis of the terms "freedom", "causation" and "responsibility"¹²⁸ in order to discover the logical functions related to them. The rationale for this was the emphasis on philosophy of language, and the idea that these logical distinctions could be discovered through the conceptual analysis of language, bringing the debate about free will to an end. The outcome for this mind-set was that the debate kept perpetuating itself separate from reality of "ordinary moral life". (McKenna & Russell 2007,9.)

As an alternative to "pure" conceptual analysis, Strawson's approach called for a naturalistic outlook starting with the basics of human psychology. This shift emphasized ordinary language at the expense of a formal one. According to McKenna & Russell, the "commitments in "Freedom and Resentment" manifest his own preference for "descriptive" over "revisionary" metaphysics. " (McKenna & Russell 2007, 9-10.)

The criticism that both Strawson's naturalistic position and his descriptive take on metaphysics have inspired includes challenges from at least two different directions. Following Watson (1987),¹²⁹ McKenna & Russell

5.1.2 and 6.1, although in the article, Watson presents historical responsibility as an example of criticizing the libertarian position (Watson 1987, 279-280). This relation is further debated by various authors, and the discussion is followed here.

¹²⁸ Even though the method of this dissertation is itself a form of conceptual analysis, the difference here is the partly historical difference in the relations between truth, language and knowledge. As stated before, more than precision about any supposed "final" definitions of the concepts, the interest here is more hermeneutic than technical, to loosely follow the distinction popularized by critical theory and Jürgen Habermas.

¹²⁹ The same division applies to the debate on blame as discussed in the previous chapter (4.2). See Watson, who among other points, questions the relation between blame and moral responsibility (Watson 1987, 283).

question the essentiality of the reactive emotions as being constitutive of human nature, at least to the degree that Strawson implies, as they write:

The reactive attitudes are not permanent, inescapable features of human nature, like other more basic emotions such as fear or love. Rather they should be viewed as cultural artifacts produced in particular historical and social conditions. Reactive attitudes, the critic maintains, involve a set of socially acquired dispositions and expectations and we can well imagine more radical shifts and transformations in human relationships and attitudes than Strawson allows for. (McKenna & Russell 2007, 10.)

Furthermore, the Strawsonian conception of the emotion is questionable. In the light of recent advances in our knowledge of the emotions, Strawson's separation of the emotional and the cognitive seems too extreme. McKenna & Russell ask: "Should moral sentiments or reactive attitudes be interpreted as mere feelings, without propositional content, or should they be regarded as necessarily involving beliefs of some relevant kind?"¹³⁰ (McKenna & Russell 2007, 11.)

Another point of criticism is related to his objection to the consequentialist accounts in Strawson ([1962]/2008), which is so forceful that he regards consequentialism as completely out of the question (McKenna & Russell 2007, 10-11). This might be too strict a delineation since there have been new accounts that have incorporated consequentialist elements into responsibility¹³¹ that have more appeal than the original Schlickian, "economy of threats" type of consequentialism. The latter angle, which Strawson argues against, is what Garrath Williams discusses as "utilitarian instrumentalism". The theories included within this group put responsibility-related questions in a sort of algebraic setting, in which responsibility is akin to "paying up" for certain transgressions, and managing these situations. (G. Williams 2010, 8.)

Additionally, critics, such as John Martin Fischer, Mark Ravizza and Angela M. Smith have argued that Strawson does not take the gap between holding and being responsible into account, and this is detrimental to the theory sketched in *FR* (McKenna & Russell 2007, 12).¹³²

¹³⁰ George Sher for example has proposed a desire-belief pair as the nature of blame. This will be discussed in 4.2.

¹³¹ According to Eshleman, Fischer & Ravizza (1998) are working with one form of this. Victoria McGeer also makes a provocative case in her recent article "P. F. Strawson's Consequentialism" that Strawson's theory itself contains distinctly consequentialist aspects (McGeer 2014).

¹³² The difficulties related to the gap are described in perhaps the most detailed manner by Angela M. Smith (2007), who describes the difference between the concepts originating from the various stances of the person being judged and that of the moral judge. Strawson's theory is specifically in trouble regarding the stance of the moral judge (of holding responsible), as Strawson's moral responsibility is defined in terms of affective, reactive sentiments. It would be inappropriate, Smith

The relation between the reactive attitudes and responsibility in general is another point of contention.¹³³ Strawson seems to be of the mind that the reactive attitudes are generally a reliable measure of what is morally responsible action. We might get an improper attitude based on mistaken information about intentions or the state of mind of an agent, but according to Fischer & Ravizza there are reasons to believe that the attitudes and their objects might not have the correlation that Strawson's theory requires. "It could be, for example, that an entire community has its reactive attitudes switched on or off in the wrong way and at the wrong times" as McKenna & Russell put it. (McKenna & Russell 2007, 12.)

As stated before, Strawson's essay was just that at the beginning. It did not attempt to carve out a fully articulated theory of responsibility. One omission in the article that some critics have taken issue with is the insufficient exploration of the relation between resentment and retribution in Strawson's theory. (McKenna & Russell 2007, 13.)

If Strawson is to be defended, as Eshleman implies, even though the free will vs. determinism problem is not solved by his re-focus, it is "dissolved" in that deterministic ontology ceases to be an obstruction to the practice of holding others responsible. Strawson originally lamented that the positions mentioned above "over-intellectualised" responsibility in that the agent would need to satisfy a set of justifiable, objective requirements related to the agents themselves or their freedom of choice in the metaphysical sense. Being responsible in the Strawsonian context is not dependent on any external justification. In other words, there is no need for a separate, objective judgment about moral responsibility; determinism might as well be assumed as it appears inconsequential to moral responsibility assessments in the naturalistic context.

Eshleman concludes with remarks on how influential *FR* has been since his work is so ubiquitously cited in writing on moral responsibility. However, he also correctly assesses that the article has not received universal acceptance. Since the problem of determinism is still often associated with theories of responsibility, Strawson's arguments cannot have convinced everyone. (Eshleman 2014.) Nevertheless, his pre-eminent stature among the philosophers of responsibility is the reason why he is selected as the starting-point in this work. It would be difficult to do otherwise.

As concluding remarks on criticism on Strawson, McKenna & Russell offer a summary as follows:

Perhaps the most fundamental problem with Strawson's theory in the eyes of his critics is that he has too little to say about the extent to which our reflections concerning the (historical) origins or sources of character and conduct can inhibit—if not altogether undermine—the

claims, for a person in the position of moral judge to form his/her judgments based on emotions only. (See Smith 2007.)

¹³³ See chapter 4.3.1 on Fingarette's article "Responsibility" (1966).

(normal) operation of our reactive attitudes. (McKenna & Russell 2007, 13.)

This sentiment is seconded here and, the comparison with the earlier historical forms of the concept is made as a proposed remedial action. This comparison begins in chapter 4.3.2 in the form of Marion Smiley's criticism of the modern concept of responsibility, and continues with the account of Aristotle's moral responsibility in chapter 5. In later chapters, the most important elements of these theories, such as blame and the forms of blame mitigation are compared. Before moving on to the actual historical comparison, a few noteworthy general criticisms are examined in the remaining sections.

4.3.1 HERBERT FINGARETTE: “RESPONSIBILITY”

By writing for decades on the topic of responsibility, based on his experience in psychology, Herbert Fingarette had an opportunity to ask the right questions in his article “Responsibility” (Fingarette 1966). As a contemporary of Strawson, he could present a set of queries that contrasted the emerging consensus on what responsibility could be about. He wrote: “it has been argued that “moral responsibility” is a vacuous notion. This [...] point of view, however, is usually a conclusion derived from analyses which are based on the presumption that only action or answerability or some combination of these are of the essence.” (Fingarette 1966, 58-59.) This exemplifies the stances that focus on the moment of action and disregard the history of prior events.

The article argues against a tacit agreement on what moral responsibility refers to within the bounds of moral philosophy (Ibid. 58). Also noted in the article is the social element of blame: it is decisive what the others think, or how they react. While discussed in the Strawsonian context in terms of the reactive attitudes, Fingarette appears to have sought a broader spectrum of interaction. He discusses how:

showing how a fuller understanding of ‘moral responsibility’ takes one into quite different regions. In moving into these new regions, we shall come much closer to the continental divide of morality, what Kant called ‘the extreme limit of moral inquiry’, the question why men should will the moral law. (Fingarette 1966, 59.)

Fingarette begins by giving us the classic example of involving *children*¹³⁴ (Ibid. 60). His other example is even more striking: “What if we found some “malicious villain”, otherwise rational, who did not wish for or respect those noble qualities that Kant mentioned? Such a man would not be morally “blind”; our “malicious villain” would be very knowledgeable about moral issues.” (Ibid. 60.) This “malicious villain” refers to *the psychopath*. Fingarette’s article is followed here in that the examples of children and psychopathy are considered in the cases of the main theories examined here. These extreme instances of agency, when applied to a situation, can change the result of responsibility assessments, which would be unproblematic in normal situations. While it is true that any moral theory breaks in the case of those who function on a completely different set of principles described in the theory, such as that of the psychopath, this is not always clearly stated. Examination of a theory through an individual who is by definition out of

¹³⁴ According to Fingarette’s follower, Michael S. Pritchard, who writes in *On Becoming Responsible* that “Fingarette claims that people are not morally accountable for what they do until they show signs of genuine acceptance of responsibility”; there is a certain level of maturity required for moral agency and moral responsibility as its extension (see Pritchard 1991, 45).

bounds of the entire system (of morality) might and should be pursued, in order to find out what responsibility can mean, and what the limits of the concept are.

4.3.2 MARION SMILEY: MORAL RESPONSIBILITY AND THE BOUNDARIES OF COMMUNITY

Another illuminating account of the problems of responsibility is Marion Smiley's *Moral Responsibility and the Boundaries of Community* (Smiley 1992). Smiley distinguishes three different concepts of responsibility that are useful to know, especially if the historical theories are compared with those that are in use today. The first is the classical concept of responsibility dating back to ancient Greece. The second is the Christian concept,¹³⁵ which is not pursued here other than by the comparison with the third concept: the modern concept of responsibility. (Ibid. 73.)

The crucial difference between the classical and the modern concept here is the location of praise or blame in the process of responsibility assessment. The modern conception, as Strawson exemplified, importantly views praise or blame as definitive of moral responsibility, but the interest in it stops there. Praise or blame as the reactive sentiment is the end state of responsibility assessment so far as the modern concept is concerned. Aristotle's theory is such that the classical concept views praise and blame more as actual, social links between the agent and the society that he or she inhabits. Therefore, responsibility assessment in the classical concept does not stop when it is clear whether the agent receives blame or not, but rather asks further where the blame is directed and of what magnitude the emotion is. Whether an individual is responsible is clear only after the big picture is understood.

By focusing her attention solely on the concept of moral responsibility rather than diverting attention toward the notions of free will and determinism, Smiley's observations produce interesting results. She describes the differences between the classical concept and the modern concept thus. There is a jarring contrast between the way academic philosophers are content to discuss moral responsibility and the actual problems of the world (see *ibid.*, 1-2). Her claim is that moral responsibility at least in the modern context is firmly rooted within social and political practice and not as some outside and superior ideal form. What Smiley wants to emphasize is that our practical judgments would be opened up in terms of moral responsibility, and that *that* understanding could help us in real situations involving the suffering of individuals and other cases of injustice.

¹³⁵ Smiley goes through the Christian concept of responsibility relatively briefly in her chapter 3, stating that while a unifying generalization of the Christian concept would be misleading because of the scholastic debates that took place during the middle ages, it is not impossible to draw conclusions about some of the common clauses that the medieval scholars agreed upon, one of these being God's role as the supreme *blamer*. While not necessarily relevant in context of this dissertation, the Christian concept of responsibility may constitute a good entry point into topics discussed here, as Smiley makes the comparison between the concepts of *sin* and *moral responsibility*. The latter concept can be described from the Christian perspective as *sin*, without the authority of God. (See Smiley 1992, 58-71.)

(See *ibid.*, 14.) She writes that “the very pervasiveness of the modern concept of moral responsibility makes it difficult to articulate.” (*Ibid.* 73.) She argues that the modern concept shares the ambiguity of its predecessors, the classical and Christian concepts of responsibility. Nevertheless, she also argues that these involve a “cluster of shared assumptions” that allow contemporary individuals to understand each other in communications related to moral responsibility. The trouble is however that even as the content of these assumptions is generally the same between everyone, the exact composition of these aggregate terms is not identical with everyone. This is why Smiley insists that the concepts of responsibility are difficult to present as one overarching concept. (*Ibid.*)

According to Smiley the modern concept of moral responsibility is not supposed to compete with the previous definitions of responsibility as a *concept* per se. Rather, what Smiley demands from the modern conception is that it strive to describe something *real*.¹³⁶ Moral responsibility would thus not only be more tangible than its secular counterparts, but also a distinct layer of morality, not reducible to the other historical definitions of moral responsibility. (*Ibid.* 73-74.)

Smiley quotes Frankena,¹³⁷ who wrote a basic introduction, *Ethics* (1973), to the questions on responsibility that were relevant for most of the last decade. She follows Frankena in the sense that he “captures the conflation of causation and blameworthiness that is characteristic of our modern concept of moral responsibility” (Smiley 1992, 74). Another source of inspiration is Aristotle’s theory: from her perspective the responsibility for action as described in *NE* III.1, including the notions of causation and voluntariness, are seen as one of the two primary conditions of blameworthiness. The other is that action is done in knowledge. (*Ibid.* 37-44.) Smiley’s contribution is most valued here for the effort to clarify the basic components and the differences between the three historical concepts of responsibility. After making these clarifications, she sets out to describe an alternative way of conceptualizing responsibility, which emphasizes a thing called contra-causal freedom.

A few more words should be devoted to the conflation between causal responsibility and blameworthiness. As Smiley very clearly puts it in the conclusion of her book, the modern concept of moral responsibility brings together the notions of causal responsibility for harm and the fact that the

¹³⁶Smiley quotes J. Feinberg’s question in *Doing and Deserving* (1970, 30) of whether the defendant is *really* responsible. A separate idea of “real” theoretical responsibility that is distinct from practical, legal responsibility is typical of the accounts of moral philosophy. The concepts of “*moral* obligation”, “*moral* guilt” and “*moral* responsibility” are used to express this emphasis (Feinberg 1970, 30).

¹³⁷ Frankena writes in *Ethics* about the distinct nature of causation and blameworthiness. (Frankena 1963, 56.)

agent ascribed that property is blameworthy. She argues further that neither of these notions is purely factual, but that they require the social dimension in order to make sense. (Smiley 1992, 255.)

Without going much deeper into the other subject, the most important distinctions from Smiley are followed here and used in further passages. These are: 1. the distinction between the classical, Christian and modern concepts and 2. the conflation between causal responsibility and blameworthiness, which she says is typical of the modern concept. What is also emphasized below is 3. the abstraction of the community, i.e., where the origin of the moral sentiments of praise and blame in the case of the modern concept of responsibility is postulated instead of actually being required to take place.

She concludes by stating that further use and understanding of moral responsibility judgments require the disputants to become conscious of the communal borders and everything that this entails in terms of expectations of each other as well as social roles and power relations within that society.¹³⁸ All in all, Smiley's book is very interesting because of its helpful distinction in re-evaluating the nature and use of moral responsibility, and it is also a very brave take on the most common assumptions that the concept has accrued within the context of the philosophy of responsibility in the previous century. (Smiley 1992, 258.)

¹³⁸ This is something that is relevant in the case of the "standing to blame" topic in ch. 4.3.4 as well as below in the case of John Doris's work, which is discussed in 6.1.

4.3.3 BERNARD WILLIAMS: *SHAME AND NECESSITY* & “MORAL RESPONSIBILITY AND POLITICAL FREEDOM”

As was noted in connection with Wallace’s rendition of Strawsonian responsibility, intention is a factor that cannot be disregarded when responsibility assessments are made. Bernard Williams, an esteemed philosopher who had discussed the specifics of theories related to responsibility had already been influential for decades. His works *Shame and Necessity* (1993) and “Moral Responsibility and Political Freedom” (1997) are important in that they offer a distinction that is useful to apply to the various theories of responsibility.

The work of Bernard Williams on the concept of responsibility will be looked at after considering how the historical concepts of responsibility can help understand the concept. The classical and modern concepts discussed before do have something in common at least in structure. As for style, Williams is also refreshingly clear-headed about the problem concerning “freedom of the will”. He recounts in *Shame and Necessity* (1993) that:

[t]he problem [...] exists only for those who have metaphysical expectations. Just as there is a “problem of evil” only for those who expect the world to be good, there is a problem of free will only for those who think that the notion of the voluntary can be metaphysically deepened. In truth, though it may be extended or contracted in various ways, it can hardly be deepened at all. What threatens it is the attempt to make it profound, and the effect of trying to deepen it is to put it beyond all recognition. The Greeks were not involved in those attempts; this is one of the places at which we encounter their gift for being superficial out of profundity. (B. Williams 1993, 68.)

This line of argument is continued in “Moral Responsibility and Political Freedom” (1997) with the following:

Moral responsibility has a function, and there is much to be said for its doing the work of ascriptions of responsibility and dealing with our responses [in form of praise and blame] to offences, so long as one does not take it too seriously.^[139] But if one accepts the idea of moral responsibility, this is often taken to imply that moral responsibility is self-applying and does not need or permit any external justification or assessment; that it is profound, or can be made so; and that it is the ideal in terms of which institutions are to be judged and to which they should try to approximate. All of these further conceptions are false. If moral responsibility necessarily

¹³⁹ A supposition here is that the hard determinist position is motivated precisely by taking moral responsibility with absolute seriousness.

involved these conceptions, it would have to be abandoned. (B. Williams 1997, 102.)

These points lead to Williams' view of moral responsibility, which is interpreted here to be a step toward a more pragmatic account of the term.¹⁴⁰ He writes about the futility of trying to put together a unified theory of responsibility by comparing the modern conception of moral responsibility with that held by inhabitants of ancient Greece as well as providing a set of elements of responsibility to illustrate these ideas more clearly (See B. Williams 1993, 55). These ideas are examined below in greater detail.

He compares our conception of responsibility with that of the Greeks and notes how the Greeks would use different conceptions for it, depending on the situation.¹⁴¹ What Williams's is saying is that – ultimately – there can be no universally acceptable definition or a concept of responsibility, as so much depends on particular circumstances. If considered in his own concept-pair of *thick* and *thin* concepts, he is of the mind that the latter is possible and useful, but attempts to arrive at the former will be met with insurmountable difficulties. As Williams's worries are shared here, his solution is also taken note of. He argues that all theories that discuss responsibility put an emphasis on one or more of the four following elements of responsibility:

- *cause*,
- *intention*,
- *state* and
- *response*.

(Williams 1993, chapter 3, & 1997, 96.)

Williams claims that these four elements, common to all attempts to classify responsibility, can be distinguished and found in those theories to varying degrees, to the extent that they are already in place in the texts of Homer. He justifies the uses of the elements instead, defining some version of an all-purpose concept of responsibility as follows:

There is not, and there never could be, just one appropriate way of adjusting these elements to one another – as we might put it, just one

¹⁴⁰ Smiley's work (1992) can be read as a continuation of this trend.

¹⁴¹ For an example, Williams refers to the *Odyssey*, where Telemachus opens the door to the armory, and lets the enemies of Ulysses arm themselves. The son claims responsibility for the event and by using the word *aitia*. Nevertheless in Aristotle's case his responsibility-related thoughts based on the word *telos* are equally important. The difference between these concepts will be briefly explored in the next section.

correct conception of responsibility. Quite apart from the differences between our practices and those of the Greeks, we ourselves, in various circumstances, need different conceptions of it. All the conceptions of it are constructed by interpreting in different ways these four elements and varying the emphasis between them. The four elements are already in Homer, and they would have to be, since the need for them, and for ideas that bring them together in some pattern or another, follows simply from some universal banalities. Everywhere, human beings act, and their actions cause things to happen, and sometimes they intend those things, and sometimes they do not; everywhere, what is brought about is sometimes to be regretted or deplored, by the agent or by others who suffer from it or by both; Wherever all this is possible, there must be some interest in the agent's intentions, if only to understand what has happened; [...] Again it must be a possible question how the intentions and actions of an agent at a given time fit in with, or fail to fit in with, his intentions and actions at other times. Under any social circumstances at all, that is a question for other people who have to live with him.

These really are universal materials. What we must not suppose is that they are always related to one another in the same way or, indeed, that there is one ideal way in which they should be related to one another. There are many ways of relating them, in particular of relating intention and state to response. (B. Williams 1993, 55-56.)

Because these elements are rarely dealt with identically or even in equal measure between the theories, Williams remains skeptical about a shared understanding of the theme between theorists. Nevertheless, from the viewpoint of this study, his elements of responsibility are, indeed, very useful. To summarize what has been the focus of the research on responsibility broadly, the cause and the intention parts have been it in the past, the response element is the popular topic now, but what is attempted here is to highlight the state element and how it appears in these different theories.

Williams sums the elements up in his article "Moral Responsibility and Political Freedom" (B. Williams 1997) by writing that "[A]ny conception of responsibility involves four elements: cause, intention, state and response" (ibid. 96). In this case, cause refers to any concrete actions taken by agents and the causal responsibility related to them. This is essentially the same concept that Coates & Tognazzini alluded to in their "causal or explanatory responsibility" and its component acts. By itself, the element of cause has not held the interest of the recent moral philosophical inquiries. (Coates & Tognazzini 2014.) Williams maintains that there can be no responsibility without causes and that only a possible case of a responsibility related act without a cause would be some sort of "scapegoat" situation (B. Williams 1993, 56).

Intention has attracted a lot more attention in many schools of thought. There is a difference, for example, in how intention is discussed in the continental circles than within the Anglo-analytical tradition. Intention in the Anglophone analytic ethics sense is more closely related to the acts. Williams explains the notion of the intention in *Shame and Necessity* (1993), in how it appeared in ancient Greece through the Homeric example of Telemachus leaving the door to the armory open, allowing Ulysses' enemies access to the weapons. At the simplest level, it is the condition of the actor in which the act is done voluntarily. (See Williams 1993, 66.)

Intentional action does not, at least according to Williams, indicate that the act is done in a normal state of mind. The element of state is a description of the effect the state of mind has on the act and the responsibility assessment. There have been examples as early as Homer of intentional acts that were not done in a normal state of mind. Williams gives the example of Agamemnon, where the king explains a situation in which the actor is not morally responsible (*aitios*), because of external influences (the gods) having an effect on his conduct. (Ibid. 52.)

In the later article, Williams further elaborates the elements of intention and state. Together these elements constitute the voluntariness of the action, as it is discussed for example in Aristotle's theory. When both of these requirements are true for agent A doing X, as Williams puts it "A does X intentionally in a normal state of mind." (Williams 1997, 97.) According to him, intention describes the willing of what the agent himself or herself "means to do", while state refers to the "state of mind" of the agent at the moment of intending and doing act X. It is not always easy to differentiate between these two elements. Williams acknowledges this issue: "There is great indeterminacy and vagueness about the agent's psychological state and the soundness of associated [reasons ...] lying between state and intention." (Ibid. 101.)

As for the final element of *response*, Williams describes it as something that "is expected, demanded, or required of the agent or is imposed on him." He notes that in the criminal law context, response usually refers only to punishment, but that in the broader context of moral philosophy there are other responses related to responsibility. (Ibid. 1997, 98.) As the content of criminal law is not under examination here, response, in terms of moral responsibility related to conduct basically just refers to praise or blame, or to a set of reactive sentiments such as Strawson's resentment, indignation and guilt.

Williams did not take the explication into very great detail: What is drawn upon here is only two sources where the elements are mentioned. The first of these is a book that discusses – among other topics – ancient theories of responsibility and the second is a relatively brief article arguing in a more political arena. In any case, Williams's explication seems to be a useful distinction that is slightly more complex than the often paired distinctions

that are usually employed, such as accountability/attributability and prospective/retrospective responsibilities, etc.

In a similar vein, Fingarette began his article by criticizing his contemporary argument on moral responsibility for focussing merely on the elements of responsibility and not the entire concept (Fingarette 1966, 58). Of course, Bernard Williams wrote decades after Fingarette, so his distinctions were not the target of the article. Williams could be exonerated from Fingarette's critique on the grounds that the distinctions Williams makes with his elements are taking the discussion of moral responsibility forwards, and that is more than the tacit acknowledgement of the concept at Fingarette's time.

A more extensive elaboration on forms of responsibility is attempted by John M. Fischer and Mark Ravizza in their *Responsibility and Control* (1998). While not directly used here, it is examined next as having some interesting connections to the other theories discussed here.¹⁴²

Williams is not alone in attempting to formulate this sort of taxonomy. Fisher & Ravizza present a similar idea. Their concept of reasons-responsiveness describes the condition in which William's state element is true. In such a case, the agent is viewed as an adult, fully capable member of the moral community. Fischer describes the condition of reasons-responsiveness in the earlier article "Responsiveness and moral responsibility" in Schoeman (1987): "An agent is morally responsible for an action insofar as he is rationally accessible to certain kinds of attitudes and activities as a result of performing the action." (Fischer 1987, 81.) Also, "a weakly responsive mechanism is all that is required for moral responsibility. [...] Moral responsibility requires *some* connection between reason and action, but the fit can be quite loose." (Ibid, 89-90.)

In *Responsibility and Control: A Theory of Moral Responsibility* (1998) Fischer & Ravizza explicitly select the Strawsonian view of moral responsibility as the basis of their theory.¹⁴³ The reason why this is so is their belief that Strawson's theory accurately reflects the implicit cultural values held by them (see Fischer & Ravizza 1998, 8-10). They follow the assumption that responsibility requires an agent who is an applicable target of the reactive sentiments (see Fischer & Ravizza 1998, 241-242). They also discuss Aristotle's primary influence on the theories of responsibility through the two excusing conditions of "ignorance" and "force" (Fischer & Ravizza 1998, 12). Fischer and Ravizza's taxonomy for the levels of responsiveness corresponds with moral responsibility in terms of whether the act is deemed blameworthy. Causal responsibility is given, otherwise it would not make sense to evaluate the act correctly. The taxonomy begins with acts that are

¹⁴² Some points from Fischer and Ravizza's *Responsibility and Control* (1998) are taken into consideration, including the taxonomy of reasons-responsive responsibilities (Fischer & Ravizza 1998, 82). Fischer and Ravizza, along with Gary Watson, are two of the few authors who discuss a facet of responsibility called *historical responsibility*.

¹⁴³ "On the Strawsonian view we are adopting in this book..." (Fischer & Ravizza 1998, 8).

mere bodily movements.¹⁴⁴ Some of these are unintentional, while others are intentional. (Fischer & Ravizza 1998, 83.)

The taxonomy-tree in Fischer and Ravizza (1998) follows the intentional bodily movements, dividing them into reasons-responsive acts and unresponsive, intentional acts. As an example of these, the writers describe a skipper who is panicked in a storm. As a result of diminished reasons-responsiveness due to external coercive factors, the skipper is not morally responsible for her actions. The writers place “insane persons” who can be determined to act out of intention at the same level of reasons-responsiveness, but their “mental illness is so severe that he acts on a mechanism that is not at all responsive to reasons.” (Ibid. 82.)

The actors on the “Responsive to reason” level of the taxonomy are divided into classes which distinguish moral and non-moral reasons. Those responsive only to non-moral reasons are either “babies” or “certain psychopaths” according to Fischer and Ravizza. This group is the same as Strawson’s second sub-group, or the group of “exempted” agents described by Watson and Wallace, examples which are examined later in greater detail. Finally, the most reasons-responsive group is divided into cases in which significant duress may or may not be present. For example, a calm skipper can remain reasons-responsive even under duress (ibid. 83). The command to abandon ship can be done rationally, and the skipper be appropriately held responsible for it. Compared to Williams’ elements, both intention and state elements are without fault in such cases. Depending on the outcome praise or blame might be relevant.

The final point of interest in this case is the thoughts that Fischer and Ravizza spend on the nature of historicity of facts, and the effect of these on moral responsibility. They make an important distinction in the chapter devoted to examination of the historical qualities of responsibility. The division of phenomena into historical and non-historical, or “current time-slice” phenomena is relevant. The historical phenomena are in some way dependent on their history, or events in the past that have affected them, while those that are non-historical are apparent without this historical knowledge; for example, we can perceive the color of an object without any knowledge about how it became so. A contrasting example Fischer and Ravizza give is a painting by Picasso, the genuineness of which cannot be determined on the “current time-slice” context. (Fischer & Ravizza 1998, 171-172.)

Fischer and Ravizza further compare Robert Nozick’s critique of distributive justice with that of historical phenomena and responsibility. Just as Nozick’s critique of justice demands that it is not enough to look at “current time-slice” situations in order to determine the justice of any given distribution of property, it is important in cases of moral responsibility to

take the historical phenomena into account as well. (See Fischer & Ravizza 1998, 178-182.) The authors further extend the considerations of the historical component of responsibility¹⁴⁵ in order to criticise Frankfurt's theory which, they argue, adheres to the "current time-slice notion". Fischer and Ravizza conclude with the statement that "Frankfurt is *correct* in thinking that taking responsibility is crucial to moral responsibility, but he is *wrong* to think that taking responsibility can be analyzed simply by reference to snapshot properties." (Ibid. 199-200.) The proposed solution to the problem that the current time-slice theories face is that a historical component be included in the *taking responsibility* theme that the writers propose.

The conclusion of their book reads: "we suggested that moral responsibility is in the class of genuinely historical phenomena". They also take note of the possible objections of entitlement-based views and Kripkean soft facts. (Ibid. 201-206.) While Fischer & Ravizza have given thought to the taxonomy and illustrate finely in their book, and the general notion of historical component of responsibility is followed here, disagreements in other areas prevent further adherence to Fischer & Ravizza (1998): for instance, while the seemingly simple progression from bodily movements toward reasons-responsive, fully responsible acts in their taxonomy is helpful, it can be limiting when all possible cases requiring assessment of responsibility are considered. Relying solely on the taxonomy, for instance, all acts would require bodily movement as the origin of causal responsibility. It could also be imagined, say, that the taxonomy could be applied to speech-acts as well, where a random utterance might replace bodily movement, or, if classified as outside of group of bodily movements, an act could be initiated with some form of electro-magnetic controller that reads the activity of a brain, for instance.¹⁴⁶

While Fischer and Ravizza's theory is novel and offers many important insights, they will be referred only as needed from this point on. Williams's elements, however, are considered essential and will be returned to in the analysis part, as they will serve as categories of explication of Strawsonian and Aristotelian concepts of character and excuses in responsibility. Gary Watson's input is looked at as a final but important source of criticism of the Strawsonian position.

¹⁴⁵ The historical component of responsibility (discussed in the next chapter) refers to the idea first mentioned in Watson's article (1987) that the developmental history of the agent influential at the moment when responsibility for an act is assessed. Fischer & Ravizza's term "current time-slice" refers to those accounts which deny or otherwise circumvent this effect.

¹⁴⁶ Another consideration is the format of Fischer and Ravizza's theory: the locus of responsibility is different than the Strawsonian cases. For this reason, Fischer and Ravizza's shift from an agent-based approach to a mechanism-based one has been criticized by Wallace (1997), Gary Watson (2004/[2001]), and McKenna (2013).

4.3.4 GARY WATSON “RESPONSIBILITY AND THE LIMITS OF EVIL: VARIATIONS ON A STRAWSONIAN THEME”

While Strawson's “Freedom and Resentment” (2008/[1962]) may be the most influential article concerning the philosophy of responsibility, Gary Watson's commentary “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme” (1987) can be read as a companion to it.¹⁴⁷ As it is, it contains many features related to moral responsibility that are left out of the discussion by Strawson's relatively brief article. The issues and definitions introduced in that paper elevate it to one of the most important commentaries on the subject.¹⁴⁸ A few of these are highlighted here: *historical responsibility* is an especially important theme that Watson considers and that is left out of Strawson's work. *Moral address* is another theme, which starts its own line of inquiry. In addition to these, the clarifications and distinctions that Watson makes are important in advancing the discussion about moral responsibility.

In “Responsibility and the Limits of Evil” (1987) Watson describes Strawson's theory as “the expressive theory of moral responsibility”. According to Watson, it is a compatibilist theory that is not consequentialist by nature. Someone being blamed usually means for Watson the attribution of a moral fault, shortcoming, vice or a character defect. In these cases, what is conceived as responsibility is the conditions of propriety in making moral judgments. In Strawson's theory, these judgments are not sufficient by themselves for determining moral responsibility. What is left out of such a definition is the practice of holding responsible. As Watson puts it: “In a Strawsonian view, blaming is not merely fault-finding appraisal, [...] but a range of responses to the agent on the basis of such appraisals. [n10] These non-propositional responses are constitutive of the practice of holding responsible. (Watson 1987, 261-262.)

In his article Watson gives the example of Robert Harris, an individual who is “brutal, vicious, heartless, mean” (Ibid. 268-271, quote from 274). The actual story is vivid and lends credibility to Harris having the traits mentioned above. The key question that Watson asks is: “Can we be in a moral community with those who reject the basic terms of moral community”? (Ibid. 268). The implied worry is conceptualizing responsibility in a way that leaves out of the equation a significant proportion of agents who can still do wrong and be causally responsible for harm.

¹⁴⁷ “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme” was originally published in Schoeman 1987. Except for the most recent article (Watson 2014) the important contributions by Watson, including Watson 1987, are also in the compilation *Agency and Answerability* (Watson, 2004).

¹⁴⁸ Compare this with the brief bibliography of the first volume of the new series of *Oxford Studies on Agency and Responsibility*, in which Watson's article is one of six mentioned (and two are Shoemaker's own publications) (Shoemaker 2013, 11).

Watson, who has influenced the topic of history's effect on responsibility in Fischer & Ravizza (1998), is even more explicit on the subject, as the following quote from "Responsibility and the Limits of Evil" shows:

I have been exploring some ways in which the expressive theory might explain the relevance of certain historical considerations. Whatever the best explanation may be, the remarkable fact is that we are, for the most part, quite ignorant of these considerations. Why does our ignorance not give us more pause? If, for whatever reason, reactive attitudes are sensitive to historical considerations, as Strawson acknowledges, and we are largely ignorant of these matters, then it would seem that most of our reactive attitudes are hasty, perhaps even benighted, as skeptics have long maintained. (Watson 1987, 282-283.)

Watson's point is that, instead of just presenting the atrocities Watson provides the biographical history of Harris, which explains the state that the man is in through past events (Watson 1987, 272-274). He argues that the example gives our reactive sentiments a pause. This is all part of pointed criticism of Strawson's theory. According to Watson, the examples as they are related do not necessarily lead to a deterministic account of the events, but he does note the effect that the additional information has on the reactive attitudes. Knowing what the formative factors for the wrong-doer were does not cancel out the initial recoil from the atrocities, mixing the emotions instead to produce "ambivalence". "No wonder!" claims the author (ibid. 275). Watson questions the need to group historical explanations with determinism, writing that "the ways in which reactive attitudes are affected by sympathy and moral are intelligible without appealing to any of the conceptions of responsibility that Strawson eschews. [...]. Horrid backgrounds do not inevitably give rise to horrid people." (Ibid. 277.)

Assuming that determinism is true, Watson emphasizes the need to recognize the necessity of *moral luck*¹⁴⁹. As evil, in Watson's words, is caused in determined environments together by "nature and nurture", the difference between those causing it and those who are not is dependent on moral luck. Were the situation otherwise, every individual exposed to similar circumstances would act in an identical way as a result. The acceptance of this fact, according to Watson, has an unalterable effect on our judgments of responsibility. In this way, determinism has an effect on our reactive attitudes, contrary to Strawson's claims. (Ibid. 278.)

The Harris example is meant to demonstrate, by telling the story of a morally unlucky individual, that because it is impossible to be responsible for one's own formative circumstances, it must be the agent's response to those circumstances that determines whether one becomes evil or not (ibid. 279).

¹⁴⁹ Watson talks of sympathy and moral luck in his "Reasons and Responsibility" (2004/[2001]) as well. He continues the debate that Fischer & Ravizza took part in in their *Responsibility and Control*.

Watson considers that this need for authority in the face of a decision may be a powerful argument against determinism; however, he sticks to the compatibilist context.¹⁵⁰ On the other hand, he is asking about the feasibility of a type of consideration of responsibility usually reserved for incompatibilism. Giving partial consent to the libertarian term allows the type of historical content to influence the responsibility-related considerations, but by distancing himself from incompatibilist issues, he sketches out a form of responsibility that is both compatibilist, i.e., not dependent on the factuality of determinism, as with Strawson's, and open to at least some sort of influence from the historical side, as Aristotle's theory is if interpreted such that character matters for moral responsibility.

Concerning the Harris example Watson asks:

Are we [...] to suppose that at some earlier stage Harris slipped heedlessly or recklessly into patterns of thought and action which he ought to have known would eventuate in an evil character? (This seems to have been Aristotle's view in Nicomachean Ethics, Book III.5.) In that case, we would be tracing his present ways to the much less egregious faults of negligence. (Watson 1987, 280. The passage of Aristotle that Watson mentions is discussed in chapter 5.1.2.)

Here Watson runs into a problem, which could explain his separation of *excuses* and *exemptions*. In Harris's case, the requirements of treating him as a morally responsible agent are not met, and as a result, any attempt to do so is bound to fail. In the end, Watson concludes his commentary on Strawson's theory of responsibility with consideration of the way character is related to responsibility:

What we are responsible for are the particular things we consent to. We need not consider whether we are responsible for the genesis of the entity whose characteristics necessitate those acts of consent, for there is no such entity. In a way, of course, one is derivatively responsible for one's self, since one's moral self is constituted by the character of what one consents to, and one is responsible for what one consents to. (Watson 1987, 281).

¹⁵⁰ "We should be struck here by the a priori character of libertarian convictions. How is Harris's consent to be construed, and why *must* it have occurred? What evidence is there that it occurred? Why couldn't Harris just have become that way? What is the difference between his *having acquiesced to what he became* and his *simply having become that way*? The libertarian faces the following difficulty: *If there is no such difference, then the view is vacuous, for consent was supposed to explain his becoming that way.* If there is a difference, what evidence is there that it obtains in a particular case? Isn't there room for considerable doubt about this, and shouldn't libertarians, insofar as we are libertarians, be very doubtful about Harris's responsibility." (Watson 1987, 280. My emphasis.)

Watson considers the possibility that the libertarian position could be more plausible *if* the libertarians could agree with the idea that “we are responsible for what we consent to”. He states that: “These claims are far from self-evident. But they hardly amount to a ‘panicky metaphysics’” again referencing Strawson. Ultimately Watson remains skeptical about the possibility of a libertarian theory of responsibility that was not incompatibilist, although the ideas he initiated in “Responsibility and the Limits of Evil” do find traction at least in the theories of Fischer & Ravizza later on. (Watson 1987, 282.)

The division of different types of excuses into excuses and exemptions comes from the explication Watson undertakes based on Strawson’s original division of the special considerations into two different groups. Watson initially labels these two groups as type-1 and type-2 excuses but, as the examination advances, he starts to refer the type-2 excuses as having exempting conditions.¹⁵¹

In the Harris example, the prisoner, who is according to Watson “an archetypal candidate for blame” (Watson 1987, 271), refuses dialogue and by doing so casts himself out of the moral community. By demonstrating understanding of morality, utilizing it in a perverse manner, Harris sets himself apart from society. Watson thinks that Harris’s behavior is different from how a child or even a psychopath would act. This complete removal of oneself from the boundaries of responsibility leads Watson to distinguish exemptions from excuses: “Hence, if we are to appeal to the constraints of moral address to explain certain [...] type-2 pleas, we must not include among these constraints comembership in the moral community or the significant possibility of dialogue – unless, that is, evil is to be its own exemption.” (Ibid. 271.)

The way that the type-2 exemptions inhibit or eliminate responsibility, as with a variety of agents extending from children to extreme examples such as Harris, is insufficiently explained in *FR*. Watson notes this and offers his suspicion on the problem with the comment that “what might be necessary to complete [Strawson’s theory] will undermine the theory” (Watson 1987, 263). As a partial solution, Watson offers the interpretation of exemptions (type-2 excuses) and their conditions as indications of constraints on *moral address*¹⁵² (ibid.).

Watson also comments on Strawson’s naturalistic turn by looking at the relation between the emotional and the rational in the reactive attitudes. Even as Strawson is adamant about the affective nature of the reactive attitudes, Watson points out that they “are not mere effusions of feeling,

¹⁵¹ R. J. Wallace adopts the distinction later in his *Responsibility and the Moral Emotions* (1994). Wallace is discussed above, in chapter 4.1.

¹⁵² See Garrath Williams’s description on the topic of moral address in his bibliography on “Moral Responsibility” (2010).

unaffected by facts[... in] that they neither require nor permit a ‘rational justification’ of some general sort. [...] Reactive attitudes do have internal criteria, since they are reactions to the moral qualities exemplified by an individual’s attitudes and conduct” (Watson 1987, 259). Watson himself is in favor of a rational, judgment-based view of blame and responsibility, as he has actively asserted recently (see Watson 2013 in Coates & Tognazzini (ed.) 2013).

Watson includes an important footnote on how Strawson’s theory goes beyond the spectator theories of the 18th century British empiricists: “Contrary to some of Strawson’s discussion, responsibility does not concern only other-regarding attitudes. You can hold up yourself responsible for failing to live up to an ideal that has no particular bearing on the interests or feelings of others. It may be said that others cannot *blame* you for this failure; but that would be a moral claim.” (Watson 1987, 259, n4.)

Coates & Tognazzini consider Watson’s Harris example in their *Stanford Encyclopedia of Philosophy* article “Blame”, written in 2014. They view the article from the perspective of the concept of *moral luck*¹⁵³. In the article, Harris is described as a cruel murderer who is obviously an easy target for blame. As Strawson’s theory views blame in particular as the tell-tale signal of moral responsibility, the culpability of Harris seems inevitable. Yet, when Watson recounts the personal history of Harris, the author demonstrates that in order for Strawson’s reactive sentiments to be the constituents of moral responsibility, they are susceptible to alteration by circumstantial knowledge.¹⁵⁴ Coates & Tognazzini claim that Watson’s example demonstrates considerations of moral luck in that it asks us to imagine a situation in which we are in Harris’s shoes instead. Given the circumstances, the empirical evidence seems to tell us that each one of us could commit the worst atrocities. The fact that we do not might merely be a question of lucky coincidence.

Coates & Tognazzini point out that the “thought that one’s moral self is such a fragile thing” causes discomfort. Becoming a wrong-doer as Harris is considered in the article to various degrees, the authors labeling this concern as “a worry about *subjunctive hypocrisy*”.¹⁵⁵ Looking at the example through

¹⁵³ Moral luck is a concept that is popularized by Bernard Williams (see, for example, Williams 1981). This is explored further in connection with moral responsibility in the various articles by Gary Watson (see Watson 2004).

¹⁵⁴ Thus “we are unable to command an overall view of his life that permits the reactive attitudes to be sustained without ambivalence. That is because the biography forces us to see him as a *victim*, and so seeing him does not sit well with the reactive attitudes that are so strongly elicited by Harris’s character and conduct. [...] Harris both satisfies and violates the criteria of victimhood.” (Watson 1987, 275.)

¹⁵⁵ “Another problem with Strawson’s approach, according to Watson, is that by focusing on an agent’s current moral attitudes, it does not seem to show why ‘unfortunate formative circumstances’

the idea of standing to blame, it appears that the only reason why we might blame Harris, or indeed any other criminal, is that we are better endowed with moral luck. Further on, if this difference in moral luck was acknowledged, would that fact make it more difficult to take that privilege? (Coates & Tognazzini 2014.) Watson's critique provides many conclusions, including the insight that not every concern related to the reactive attitudes is important for moral responsibility. He then gives an example of a case in which *moral standing* has significance; there are instances where it is not appropriate for a person to blame another at all due to their different statuses in the moral community or due to the particular nature of their relationship with each another. (Watson 1987, 283.)

At this point it would be impossible to overstate the importance of Watson's work for the current state of the philosophy of responsibility. His important contributions include his challenges related to (1.) *aretaic appraisal*, which implies that Strawson's theory disqualifying the most vicious persons from his system of responsibility is a flaw in his theory, (2.) as well as is its incomplete development of the aspect of *historical responsibility*. Additionally, (3.) Watson's observations pave way for a lighter concept of *moral character*, that may be useful, if the thicker definitions of character prove to be unsuitable for moral theories.¹⁵⁶ Continuing the theme of (4.) *moral address*, Watson's critique of the emotion-based account of blame,¹⁵⁷ leading to the questions of *standing to blame* in the relations between the agents in the Strawsonian system. (5) The aforementioned distinctions of *accountability/attributability* and the need to take both into account in order to fully understand moral responsibility is another important contribution (6) Finally, the division between *excuses* and *exemptions* has proven useful to subsequent philosophy of responsibility.

are relevant to exempting conditions." (Schoeman (ed.) 1987, 21-22. Also noted in McKenna & Russell 2007.)

¹⁵⁶ See Schoeman's summary of Watson's article in Schoeman (ed.) 1987, 22-21.

¹⁵⁷ In his recent article in Coates & Tognazzini's *Blame* (2013) called "Standing in Judgment" Watson contrasts *nonjudgmentalism* with *judgmentalism*, emphasizing Watson's stance of blame being more judgment in nature than an emotion.

4.4 SUMMARY

The most cited contemporary Strawsonian, R. Jay Wallace, and his version of theory of responsibility was described as part of the broader picture of the Strawsonian account of responsibility. How Wallace conceived responsibility, it was noted, was by following Strawson's example closely with a couple of notable differences. The first of these was that, more explicitly than Strawson, Wallace's responsibility was associated with the idea of holding one responsible, and even whether it is fair to hold one responsible. Arguably, this divorces Wallace's theory from the actual interactions between human beings a step further, but his book *Responsibility and the Moral Sentiments* (Wallace 1994) can be read as the most carefully articulated theory expanding the Strawsonian position. The other difference had to do with Wallace's explicit Kantian element, which leads him in the end to utilize the concept of *reflective self-control*.

After Wallace's work, attention was turned toward the contemporary debate on the concept of blame. Strawson is the definitive author here as well, as his reactive sentiments based view is the influential position to which the others refer. Alongside Strawson's theory, two modifications to the default position were looked at, one by Scanlon and the other by Sher. Scanlon's concept of blame is more clearly grounded in sociality than Strawson, as it focuses on the interpersonal relations between the agents. Blame implies an unfulfilled expectation of social conduct and the resultant adjustment of attitude toward the other. Scanlon's view has attracted criticism since, including from Wallace, on the grounds that Scanlon's theory dismisses the emotional side of blame. Sher's blame on the other hand seems to intend a synthesis between the cognitive and the emotional aspects of blame, as within his theory, blame consists of belief-desire pairs.

The last substantial section (4.3) focused on the most important arguments highlighting something that the Strawsonian position, as it is understood here, misses either by design, or by accident: a fellow philosopher at Strawson's time, Herbert Fingarette, wrote in his 1966 article that responsibility as a word can refer to other content that is limited by the relatively constrained concept of moral responsibility. On the other hand, Marion Smiley's book *Moral Responsibility and the Boundaries of Community* (1992) explored other historical conceptions of responsibility leading her to suggest three alternative concepts: the classical, the Christian and the modern concept of responsibility. She claims that the modern concept has conflated the ideas of causal responsibility and blameworthiness.

Working from a similar angle (and also directly influencing Smiley with his earlier work) Bernard Williams in *Shame and Necessity* (1993) and "Moral Responsibility and Political Freedom" (1997) sketched out four elements of responsibility which, he argues, were present in each attempt to

formulate theories of moral responsibility since the ancient Greeks did their work.

Gary Watson's article "Responsibility and the Limits of Evil" (1987) was examined for numerous pointed questions that show that the Strawsonian position is inadequate as far as it is used to define responsibility. As the following chapter is devoted to another strand of history and a different conception of responsibility, the themes of the chapter above will defer to the other perspective, but selected questions will be returned to the chapters devoted to the systematic presentation of the concepts (chapter 6) and their analysis (chapter 7) in part II.

This concludes the substantive section on Strawson and the responses to his work. In the next chapter an altogether different theory of responsibility is examined. Aristotle's theory of moral responsibility is based on his work in ancient Greece, but recent scholarship has breathed new life into his ideas, many of which can be compared to those held by modern philosophers. As a point of contrast, the examination of this other theory is considered to be able to highlight important topics that are universal in human thought as well as to point out the possible omissions in the current, focused philosophical efforts.

5 ARISTOTELIAN RESPONSIBILITY¹⁵⁸

Aristotle (384 – 322 BC) was the first philosopher who, in a broad sense, formulated a theory of responsibility that is still worth examining. His account is based on his virtue ethics, which are described in detail in the following sections. Among other key concepts within the statement of this idea of moral responsibility, praise and blame were described as an appropriate response either to acts committed or to character traits. (Eshleman 2014, quoting EN 1109b30-35).

Aristotle has provided inspiration in the debate on responsibility for contemporary scholars as well, since his conceptualization of the subject is still free of the problem of determinism. Side-stepping this difficult issue makes a number of related topics seem clearer and thus more appealing to apply to real life. As an obvious difference, the problem of determinism has occupied philosophical authors of the last century (see Eshleman 2014). As Aristotle did not concern himself with the problem as much as his modern counterparts, his writings are (depending on one's viewpoint) refreshingly free from its burdens. Whether Aristotle's theory can be considered viable when subjected to the problem is an issue that has attracted arguments for and against. There seems to be a chance that Aristotle's work fits into the definition of compatibilist theory, and thus remains untouched by the problem.

What is unique about Aristotle's agency is also noteworthy: Aristotle's choice that depends on a "desire resulting from deliberation, one that expresses the agent's conception of what's good" is an important distinguishing factor (Eshleman 2014, quoting EN 1111b5-1113b3). This conscious understanding of available courses of action in choice-making situations, the type of control wielded by the agent in question, separated from the epistemic conditions of responsibility, is related to Aristotle's view on character and its effect on responsibility. Relations between Aristotelian conceptions of moral agency and character, as well as the problematic notions related to the responsibility of character are discussed in the following sections.

The transition of Aristotle's thoughts into our culture and age raises some issues. Some degree of interpretation is necessary as the ambiguity in Aristotle's theory is at odds with the current debates on responsibility as they

¹⁵⁸ This chapter is an extended version of the author's "Teot, luonne ja kansalaisuus - Nykyfilosofian tulkintoja Aristoteleen vastuukäsityksestä" in the Finnish-language publication *Historiallinen aikakauskirja* (02/2011). The article discussed three different interpretations of what "responsibility" can mean within Aristotle's works *Nicomachean Ethics* and *Politics*. Compared to the work presented here, this article provides a summary of the sources used here and a brief discussion of the questions that are discussed further below.

appear today, It is hoped that insights into the Aristotelian concept of responsibility can be gained through interpretation, following attempts by distinguished scholars.¹⁵⁹

Historically, the initial form of the discussions of responsibility is found in Aristotle's *Nicomachean Ethics* (*NE*). Aristotle's ideas influenced Christian philosophy extensively through the works of Augustine and Thomas Aquinas among others. As an ethicist, David Hume can be read as returning to the classical concept, but with the added ideas introduced by the empiricist methods. More recently, Strawson's effort has also led to renewed interest in Aristotle's concept of responsibility because it provides the modern interpretations.¹⁶⁰ These recent contributions are examined below in the order that Aristotle's theory is discussed in chapter 5.1, where it is explicated in three different interpretations: responsibility for action, responsibility for character and political responsibility. The discussion related to the most contested topic in Aristotle's responsibility, whether the idea of responsibility for character is to be included in his theory of moral responsibility, is then discussed in chapter 5.2.

¹⁵⁹ Texts by scholars forming the basis for understanding Aristotle's responsibility include (but are not limited to): Irwin 1980; Roberts 1987; Broadie 1991 & Meyer 1993.

¹⁶⁰ The differences between the classical concept of responsibility and that of modern authors were discussed in the previous chapter (4.3.2).

5.1 ARISTOTLE'S THEORY OF RESPONSIBILITY

Aristotle is relevant in terms of *moral responsibility*. The academic discussion on the concept has been present in one form or another in Western philosophy since its very beginning. There has hardly been any continuous process of seeking a universally accepted definition. For this reason, the writings on responsibility by Aristotle are still as valid as any other more recent theories. The connection between the themes of knowledge and responsibility and whether one can be blamed for acts done in ignorance are questions which still recall Aristotle, and are no less credible because of it.

In the more mainstream utilitarian and Kantian accounts of responsibility in ethics (and political philosophy) the concept has had an ancillary role to the larger wholes in the context of issues such as rights and duties. Rather, the topic has been revisited from time to time and from different viewpoints and, as we saw, the concept of responsibility has usually had a secondary role to another concept, for example, *right* or *duty*. Not until the 20th century has moral responsibility – as its own topic – received as much attention as it does now. In recent decades, research on Aristotle's conception of responsibility has been a topic of moderate interest. The turn of the 1980's and 1990's particularly saw a lot of publications devoted to the subject.

As Aristotle does not have any distinct word that could be directly translated as "responsibility", some interpretation has to be used.¹⁶¹ *Aitios/aitia* generally is considered to be the ancient Greek equivalent, but since this word does not encompass every notion of the Aristotelian concept, it is difficult to draw direct comparisons between the contemporary theories and the ancient counterpart. This does not mean that this is impossible, nor that there might not have been any attempt to do this. This brings us to the question about interpretation needed for the Aristotelian notions to work in today's context. As stated, Aristotle does not have a single word that would translate directly as "responsibility". Instead concepts such as *aitios* implying guilt and *telos* signifying the ultimate goal of the action should be taken into account, as are many other elements. The different approaches considered here produce different points of emphasis in Aristotle's context: these are (a)

¹⁶¹ By looking at interpretations derived from Aristotle's theory of responsibility, by a rough account three alternatives emerge: 1. to attempt to remain within the contemporary context with Aristotle; i.e., to try to engage with philosophy and concepts that the man himself could understand. 2. To take the scholarly route and try to learn from Aristotle and apply his teachings as appropriate to the current world, and 3. finally to take a Nussbaumian approach (see, for example, Nussbaum 1986) and build upon the Aristotelian foundations. This last strategy is commonly referred to as neo-Aristotelianism. The strategy of interpretation followed here belongs to the 2nd category. It is noted however, that some of the sources used here are derived from the third group, and are as such respected as a source of new thought.

responsibility for action, (b) responsibility for character and (c) political responsibility. Actions are significant for the first interpretation, the actor for the second, while the third looks at the previous two in relation to society. Through these three interpretations of Aristotle's responsibility, it is argued that the significant difference between the modern conception of responsibility, and the classical, Aristotelian responsibility is the inclusion of the character-based considerations in determining responsibility (see Smiley in the previous chapter 3.4.2.)

These interpretations have originated from different disciplines. As a general guideline interpretation, (a) is used by mainstream theoretical traditions of normative ethics including consequentialist and Kant-influenced, deontological theories. In addition to these theories, interpretation (b) is indebted to virtue ethics such as that of G. E. M. Anscombe.¹⁶² Interpretation (c) is derived from political theory, which is not part of moral responsibility as such directly, but in Aristotle's context has relevance for the idea. Instead of selecting a single interpretation, three different viewpoints are looked at here in an attempt to describe how diverse the discussion on responsibility today is. As we saw in the methodology section (2.1.), whenever the disciplines referred to here include more than one, a naturalistic perspective would be the fall-back discipline. In this, T. H. Irwin's 1980 article "Reason and Responsibility in Aristotle" forms the basis of discussing Aristotle's responsibility.¹⁶³ Other significant sources frequently used here are Roberts (1989), S. Broadie (1991)¹⁶⁴ and B. Williams (1993). Susan Sauvé Meyer's interpretation Meyer's *Aristotle on Moral Responsibility* (Meyer 2011, see 161-167) is "roughly naturalist" in following Strawson's conception of responsibility in general terms. This makes for an easy comparison, but it needs to be noted that the same risks of misinterpretation that Meyer notes apply here, In the same vein, Susanne

¹⁶² See G. E. M. Anscombe (1958) "Modern Moral Philosophy" and particularly Anscombe (1977) "Thought and Action in Aristotle", in J. Barnes, M. Schofield & R. Sorabji (eds.) *Articles on Aristotle 2. Ethics and Politics*, Surrey: Duckworth.

¹⁶³ Terence H. Irwin's article "Reason and Responsibility in Aristotle" in Amelie Oksenberg Rorty's *Essays on Aristotle's Ethics* (1980) summarizes Aristotle's responsibility for action, which is the subject of chapter 5.1.1. Irwin's text is perhaps the most rigorous article devoted to clarifying the exact formulations related to Aristotle's theory of responsibility. Irwin's scholarship of the ancient philosopher is accurate and interesting. Only the speculative additions to the original theory have come under scrutiny, and these extended parts, such as the "Complex Theory" have attracted more criticism, especially from Jean Roberts (1989). Additionally, Irwin draws interesting comparisons between the theories of responsibility of Aristotle and David Hume.

¹⁶⁴ The article "Responsibility for Action and Character" by Jean Roberts (1989) and the major work *Ethics with Aristotle* by Sarah Broadie (1991) are looked at on Aristotle's responsibility for character. These sources are read in connection with Meyer's book. Roberts's article originally discusses the relationship between Aristotle and his teacher Plato. One of the major emergent themes from this influential comparison is the fine details of the Aristotelian account of moral responsibility.

Bobzien has recently written a clear and helpful article in the *Cambridge Companion to Nicomachean Ethics*, which is also referred to here (Polansky (ed.) 2014).¹⁶⁵

Moral responsibility in Aristotle's context is understood as including both responsibility for action and character-based concerns. Compared with the contemporary understanding of the concept, the specific treatment of character and its influence on responsibility sets Aristotle's moral responsibility apart from them. For an example, Jean Roberts writes of these two distinct notions in her "Aristotle on Responsibility for Action and Character" (1989). Political responsibility however is a subject that is difficult to fit within the boundaries of the modern concept of moral responsibility, even as both social and political concerns are discussed in *Politics*.¹⁶⁶ As for Aristotle's moral responsibility, the most important source is *Nicomachean Ethics*¹⁶⁷ and to a lesser extent *Rhetoric*.¹⁶⁸ The relevant passages on responsibility are mostly located in the third book of the *Nicomachean Ethics*.

¹⁶⁵ Susanne Bobzien's article on Aristotle's theory of moral responsibility in *The Cambridge Companion to Aristotle's Nicomachean Ethics* as well as Javier Echeñique's book (2012) on the same subject represent the most recent commentary on the subjects discussed below. These two examples demonstrate that the book is not closed in terms of interpreting Aristotle's writings on moral responsibility: while the former would incline one to accept the comparisons that are made here, the latter argues that Aristotle's moral responsibility is incompatible with the modern concept. This debate is a persistent fixture of the Aristotle scholarship, and it will be addressed in chapter 5.2.

¹⁶⁶ Aristotle's *Politics* (Pol.) in Aristotle. 1984. *The Collected Works of Aristotle*. Ed. Jonathan Barnes. Princeton University Press, Princeton.

¹⁶⁷ Aristotle's *Nicomachean Ethics* (NE) in Aristotle 1984. *The Collected Works of Aristotle*. Ed. Jonathan Barnes. Princeton University Press, Princeton.

¹⁶⁸ Aristotle's *Rhetoric* (Rhet.) in Aristotle 1984. *The Collected Works of Aristotle*. Ed. Jonathan Barnes. Princeton University Press, Princeton.

5.1.1 RESPONSIBILITY FOR ACTION

In Aristotle's case, the conception of *responsibility for action* is used here in the sense of the most common propositions, which take the form of sentences such as "s is responsible for X", where s refers to the agent (or the actor) and X signifies any given act with moral relevance. In Aristotle's works, the most relevant passage on the subject is book III of the *Nicomachean Ethics*. Susanne Bobzien writes that, if anywhere, Aristotle's theory of moral responsibility is located in *NE* III 1-5 (Bobzien 2014, 1).¹⁶⁹ At the beginning of this passage he writes:

Since excellence is concerned with passions and actions, and on voluntary passions and actions praise and blame are bestowed, on those that are involuntary forgiveness, and sometimes also pity, to distinguish the voluntary and the involuntary is presumably necessary for those who are studying excellence and useful also for legislators with a view to the assigning both of honours and of punishments. (NE III 1, 1109b30–1109b33.)

The word itself, as we saw, lacks a direct counterpart in the ancient Greek language. Two aspects of the concept in ancient Greece can be distinguished in the writings of Aristotle, however. First, there is the more general idea of being blamed (*aitia*), which is also present in other traditional sources from ancient Greece.¹⁷⁰ Second, Aristotle scholars have until recent years emphasized voluntariness over blameworthiness. Voluntariness or *hekousion* is the second of Aristotle's two criteria for responsible action. Aristotle divides acts into voluntary and involuntary and, again, voluntary ones are divided into acts chosen and not chosen (*Rhet.* I 10, 1368b7–24). Each of the chosen acts is, he claims, done in full knowledge. The voluntariness of the act refers to the notion that action is not coerced and it is done in complete understanding (i.e., not in ignorance).¹⁷¹

¹⁶⁹ While the general consensus about Aristotle's theory of responsibility is located within *Nicomachean Ethics*, it needs to be noted that *Eudemian Ethics* (*EE*) shares three chapters with the former. The description of the details related to moral responsibility are also found in *EE*, although in a form that emphasizes voluntariness more than the counterpart in *NE*. (*EE* II.6–11.)

¹⁷⁰ In his *Shame and Necessity*, Bernard Williams uses *The Odyssey* as an example of *aitia* type of responsibility. Ulysses has already returned from his journey and is confronted with his adversaries. Telemachus has had left the door of the armoury open, enabling the enemies to arm themselves with shields and spears. Telemachus confesses that he has indeed left the door open.¹⁷⁰ In this situation he expresses that he and only he can be blamed. He expresses the intentionality of act and acceptance of possible punitive measures. (B. Williams 1993, 50–55.) For another take on *aitia*, see Susan S. Meyer's *Aristotle on Moral Responsibility* (2011, 53, n1), according to which *aitia* refers to causal responsibility only and does not have moral connotations.

¹⁷¹ The esteemed Finnish scholar and translator of the *Nicomachean Ethics* Simo Knuuttila has suggested this (Finnish edition of *NE*, p. 221 & 224–225).

An act done voluntarily and knowingly is naturally an act for which the actor is fully responsible. A coerced or otherwise involuntary act is in turn not to be blamed. An act done out of ignorance can be considered wrong in Aristotle's theory when it is voluntary. This is a part of his response to Plato's view that no one knowingly does wrong (see Roberts 1989, 23).

Responsibility for Aristotle is thus connected with voluntary acts.

Terence H. Irwin's article "Reason and Responsibility in Aristotle" (Irwin 1980) is used here in order to describe the basic conception of responsibility for action.¹⁷² Irwin describes Aristotle's responsibility not as an attributable property, but negatively as the lack of the excuses of ignorance and coercion. He argues that Aristotle's responsibility is "praised by philosophers who like to keep these things simple and unmetaphysical". (Ibid. 117.)

Irwin is in agreement that most writers on Aristotle's responsibility focus on chapters 1-5 of book III of *Nicomachean Ethics* and declares that his scope is the same. Irwin gives the theory a name on a later occasion:¹⁷³ his text distinguishes between the common interpretation of Aristotle's theory of responsibility, which he dubs "the simple theory" from "the complex theory", which contains Irwin's own additions. He writes that Aristotle's theory not only focuses on action and prominently contains study of voluntary action related to ethical and practical, jurisprudential¹⁷⁴ functions as well as the responses of praise and blame, but also some specific writings on character. Those of the latter sort will be looked at below. (Irwin 1980.)

The idea of Aristotle's responsibility for action can be briefly summarized, adhering to Irwin's formulation of Aristotle's "simple theory", as follows: "A is responsible (a proper candidate for praise and blame) for doing x if and only if A does x voluntarily" (T. H. Irwin 1980, 125). In this case, Aristotle's view of responsibility appears as retrospective, i.e., to be determined at a later point of time in relation to the act. Irwin claims that this is because, somewhat surprisingly, Aristotle does not consider forethought in connection with responsibility (ibid. 119). Thus, if the condition of voluntariness is fulfilled, the act directs praise or blame back towards the actor, being the response indicative of that for which the actor is being held responsible.

¹⁷² In the author's view, his article provides the most extensive and precise presentation of Aristotle's theory of responsibility so far.

¹⁷³ He discusses Aristotle's theory of responsibility in his encyclopedia article in *Encyclopedia of Ethics* (Becker & Becker 2001) in a brief but clear way as an "account of voluntary action and conditions for moral responsibility" (Irwin 2001, 94).

¹⁷⁴ Irwin gives an example of the actual Athenian courts and their stance on voluntary action. Apparently the courts took voluntary action to be associated with *forethought*. This notion of forethought is however insufficiently discussed in Aristotle's context, so that it remains without a clear definition. Despite this, according to Irwin, Aristotle seemed to include a possible connection with the theme, which is fascinating. This is an example of how Aristotle thought his theory was relevant in the social context of his time as well. (See Irwin 1989, 119-120.)

An early formulation of Aristotle's responsibility, which is based mainly on *Nicomachean Ethics*¹⁷⁵ but also on writings in *Eudemean Ethics* (Aristotle 2011) and *Magna Moralia*,¹⁷⁶ is described by Irwin as follows: "(5) A does x voluntarily if and only if he does x neither (a) by force nor (b) because of ignorance" (Irwin 1980, 121-123). Despite this, Irwin notes that this book contains a description of voluntariness¹⁷⁷ on the basis that: "voluntary is what we do without being compelled" (ibid. 121). Just as voluntary action is a result of desire or of a conscious decision, involuntariness, Irwin claims, implies the contrary in the sense that these are lacking from the act (ibid. 120). The two claims of (5) the requirement of action not being done out of ignorance or coercion, and that of (8) responsibility being dependent on voluntary acts; in other words, the absence of the excuses of ignorance and force as well as the combined status of being an apt candidate for praise and blame as well as acting voluntarily, together account for much of what is commonly understood about Aristotle's responsibility. Aristotle's model has since been adopted by philosophers for ages to come.¹⁷⁸

Discussing the specifics of agency, Irwin adds that: "(11) A is responsible for doing x if and only if (a) A is a normal adult, and (b) A does x voluntarily". This adds the demand for sufficient maturity of the agent in Aristotle's theory, which means that children and animals especially are excluded from considerations of moral responsibility. Irwin does however note that "to say more precisely, what a "normal adult" is, we should find some further feature distinguishing responsible agents from animals, the insane, and other voluntary non-responsible agents." (Ibid. 125.) There is some disagreement on how the group of actually responsible agents is defined in Aristotle's theory. Among other reasons for this, Irwin points out that in his context "Animals and children lack decision" (Ibid. 127). This relates to one of the most obvious reasons that Aristotle takes up the theme in *NE* III.5. Aristotle tries to set those agents that are fully mature and thus viable recipients of praise and blame apart from animals and children. Difficulties however arise from the fact that children gain the power of decision gradually, but these will be addressed later on.¹⁷⁹

¹⁷⁵ *NE* book III contains the relevant ideas (*NE* III); see also the discussion on just acts, in which Aristotle states that if the agent's action is voluntary, it is blamed (*NE* V.8).

¹⁷⁶ *Magna Moralia* is discussed less, as its authenticity as a work of Aristotle is uncertain (see Irwin 1980, 146, n9).

¹⁷⁷ In *Ethics with Aristotle* (1991), Sarah Broadie also gives an excellent treatment of the idea of voluntariness as it appears in both *Nicomachean* and *Eudemean Ethics* (see Broadie 1991, 124-178).

¹⁷⁸ The connection between responsibility and voluntariness is for example also present in Aquinas's philosophy: "St. Thomas [Aquinas ...] insists that voluntary is necessary and sufficient for responsibility" (Irwin 1980, 126).

¹⁷⁹ See Michael S. Pritchard's *On Becoming Responsible* (1991) on this theme. The situation is similar to Dennett's *plateau* metaphor. Aristotle's writings on character, which are discussed in the next chapter, fleshes out this theme in one way or another.

Aristotle's "simple theory" of responsibility represents the common understanding of his classic work, at least at the time of the publication of Irwin's article (1980), but he doesn't remain content with merely presenting the common understanding of the subject. He puts his own extended interpretation alongside it, which Irwin calls the "complex theory". Irwin introduces the concept of decision in the theory of responsibility with the claim: "(12) A is responsible for doing x if and only if (a) A has effective decision-making capacities and (b) A does x voluntarily" (ibid. 129). In this context, Irwin then raises the ambiguity of whether the decision-based requirement is based on the actual act; i.e., that the agent would have applied his capacity to "decide to do x" as is the case in Irwin's claim (14). On the other hand, the suggestion is made that it could be sufficient for the capacity to make *prohairesis*-decisions¹⁸⁰ at the time of the act, which does not necessarily have anything to do with the act in question, as is the case in claim (13). (Ibid. 131.)

Irwin thinks that the requirement for the decision (included in Irwin's claim 14) is too strong. He adds that Aristotle does not hold deciding to be necessary, and thus settles for the possibility to put the decision in the hands of the agent. The expanded interpretation, including Irwin's additional emphasis on the capacity to decide, or the "complex theory" reads as follows: "(15) A is responsible for doing x if and only if (a) A is capable of deciding effectively about x, and (b) A does x voluntarily" (Irwin 1980, 132). This limitation that the requirement for a decision brings to moral responsibility disqualifies animals, children and presumably also the mentally ill as responsible agents. The clarified definition, which is the final formulation of the "complex theory": "(15) A is responsible for doing x if and only if (a) A is capable of deciding effectively about x, and (b) A does x voluntarily". (Ibid.) In this way, it is enough for Irwin that the agent have the capacity for the decision to make her responsible for voluntary acts.

All in all, Aristotle's responsibility for action¹⁸¹ requires that the agent does X voluntarily, while lacking the excusing conditions of ignorance and force, and, according to Irwin, instead of just being a proper candidate for praise and blame, the agent has to have the capacity for a *prohairesis*-type¹⁸² decision as well. The expanded interpretation, including Irwin's additional emphasis on the capacity to make decisions, or the "complex theory", reads as follows: "(15) A is responsible for doing x if and only if (a) A is capable of deciding effectively about x, and (b) A does x voluntarily". (Irwin 1980, 132.)

¹⁸⁰ Irwin's description of "decision" is quite brief in the case of this article. Roberts argues against Irwin's interpretation (Roberts 1989, 35) Sarah Broadie puts a lot of effort into describing the concept as well (Broadie 1991, 78). See also Salkever (1992), who has additional useful descriptions for Aristotle's key concepts.

¹⁸¹ Thanks to Johanna Ahola-Launonen for an important question, which helped to clarify the relation between concepts of responsibility for action and moral agency in this context.

¹⁸² The concept of *prohairesis* is examined in the next chapter (5.1.2).

Jean Roberts' article in *Ancient Philosophy*, "Aristotle on Responsibility for Action and Character" (1989) subjects Irwin's topics to closer examination. She agrees with Irwin's initial premise and the outlines of the "simple theory" by noting that Aristotle's "voluntary actions a[re] those for which the agent is liable to praise or blame. [Aristotle] then describes involuntary action. An action is involuntary if it comes about by force [...] or because of ignorance", thus essentially describing Irwin's key points (Roberts 1989, 23). She also looks at Aristotle's work in contrast with Plato¹⁸³ in a historical context, but provides commentary on Irwin (1980) as well. Roberts argues that Irwin's complex theory simplifies Aristotle's conception of responsibility too much, and thus the claims that Irwin is making are not *actually* Aristotelian (Roberts 1989, 35, n11). This includes the notion of character, the effect of which is easily dismissed with Irwin's interpretation.

Roberts remarks that: "Character is thus not a matter of nature and is to that extent possibly voluntary." (Ibid. 30.) While Irwin distinguishes action and character as distinct themes in book III of *NE*, Roberts follows his example but uses the concept of character in a different way: by the separation of the themes of responsibility for action and character, she questions the connection between the character and voluntary action as it is understood in modern terms.¹⁸⁴ As a result, she argues against Irwin in claiming that his interpretation is not Aristotle's proper view about moral responsibility, because "it is not, as Irwin acknowledges, what Aristotle says." (Roberts 1989, 35.)

Roberts refers to a venerable cast of philosophers including Furley (1967), Williams (1985), Gauthier & Jolif (1959), Joachim (1951), Stewart (1892) as well as W. D. Ross¹⁸⁵ (1968). She quotes Ross in that she, holds Aristotle to be an incompatibilist, but as one, who did not properly examine the problem

¹⁸³ Nancy Sherman also reminds us of the different theories of virtue held by Plato and Aristotle: Plato settled for the cardinal virtues of "wisdom, courage, moderation and justice" whereas Aristotle's theory contains a significantly more detailed list of virtues. In Aristotle's theory, the capacity for the development of the virtues is required in connection with character and rational choice and deliberation. As a whole, Sherman views the Aristotelian conception of virtue as somewhere between the Stoic and Kantian theories since in its context a full life to an extent requires material fortune in addition to virtuous conduct. (Sherman 2001, 504- 505.)

¹⁸⁴ Additionally, Roberts highlights a possible misconception, which might arise when looking at the contemporary theories and Aristotle's responsibility side-by-side: if one takes for granted Aristotle's idea that we are responsible for our characters, and being so, if that person sought an alternative theory for retribution in this context, she could interpret Aristotle to mean that the "vicious chose at an earlier time to perform the sorts of actions that they wanted to become bad" when they could have chosen the opposite (Roberts 1989, 27). A person with an unjust character would not, in Aristotle's context, choose to act justly.

¹⁸⁵ Ross's view however, according to Roberts, would be misrepresenting the Aristotle's actual position, and she attempts to set the record straight by conducting an in-depth examination of Aristotle's responsibility of character.

of determinism. (Roberts 1989, 28, 35, n13.) Her findings are that: (1.) Aristotle's character is not of nature, but based on voluntary action, and thus (2.) by emphasizing character in connection with responsibility (in the form of voluntary acts) Aristotle is actually "making room for moral training and education". (Roberts 1989, 31.) While the general attitude toward Aristotle here is that he is not bound by the division, the various interpretations have an effect on how his theory is perceived. For this reason, the teleological outlook, which is often interpreted as indicating an incompatibilist stance, is looked at at the end of this chapter, although very briefly. The status quo of the naturalist, compatibilist interpretation is then returned to with reference to authors including Susan S. Meyer (2011), Myles F. Burnyeat (1999) and Nancy Sherman (1999).

Irwin argues that this view of Aristotle's concept of responsibility for action is not sufficient. Looking at other sources such as the *Eudemian Ethics* and *Magna Moralia*, he concludes that in addition to voluntariness, Aristotle would agree that responsibility requires the additional condition of a capacity to decide. Decision (*prohairesis*) is a key feature of being an ordinary adult capable of moral action in Aristotle's context. Irwin formulates this "complex theory" of Aristotle's responsibility for action as follows: "A is responsible for doing x if and only if (a) A is capable of deciding effectively about x, and (b) A does x voluntarily" (Irwin 1980, 131).

One way to approach the question would be to examine the acts through teleology. Ends have a strong role in Aristotle's philosophy, so strong in fact that (interpreted to the extreme) the ends sometimes determine the means, or the acts completing them. The importance of the ends is connected with one of the most unintuitive principles from the contemporary perspective — the principle of teleology. It is familiar in everyday thinking, but it is difficult if not impossible to accommodate to science.¹⁸⁶ Where causality is stressed nowadays and the order of instances advances in a linear temporal axis, Aristotle's case is not always as straightforward.¹⁸⁷ In his theory, the individual acts within the boundaries of nature, which means that the principles of nature affect her. The principle of teleology is one of these.

The original principle can be found in Aristotle's examination of nature, *Physics*:

Therefore action for an end is present in things which come to be and are by nature.

Further, where there is an end, all the preceding steps are for the sake of that. Now surely as in action, so in nature; and as in nature, so it is

¹⁸⁶ Compare with Stephen G. Salkever (1990) *Finding the Mean*, New Jersey: Princeton University Press, 19–21.

¹⁸⁷ While Aristotle's teleology is touched on here as a subject, it is not considered a significant enough matter, to warrant further exposition. For more in-depth looks at Aristotle's teleology, see Falcon (2012) & Pavlopoulos (2003).

in each action, if nothing interferes. Now action is for the sake of an end; therefore the nature of things also is so. (Aristotle, Physics, II 8, 199a8–199a12.)

How much the principle of teleology affects responsibility is a matter of interpretation. What kind of effect nature has on acts is essential, because according to interpretation (a), responsibility for action, voluntary acts, made in full understanding, would not have this effect at all: the principle of teleology would not be of concern. On the other hand, what nature is exactly has several different interpretations. For example, J. L. Ackrill discusses the theme of Aristotle's view of human nature in his article "Aristotle on Eudaimonia" in A. Oksenberg Rorty's *Essays on Aristotle's Ethics*. According to Ackrill, if being human included nature, the situation would again be more complex: Human nature is not precisely defined by Aristotle, leading to insufficient grounds for a thorough examination of interpretation (a) in relation to teleology. (See Ackrill 1980, 33.)

The most direct way to distinguish whether the means explain the ends or the ends explain the means is related to four types of causes used by Aristotle: material, formal, effective and final. In the case of responsibility, effective and final causes determine whether the explanation is causal (former) or teleological (latter). Even though the four causes are distinct, they can be in effect simultaneously.

In brief, responsibility according to interpretation (a) is based on the following: an act originates from the individual and is directed toward its end. When this end is actualised, or when a consequence of the act actualises this end, its effect on the society of which the individual is a part, incites feedback in form of praise or blame. If, for the sake of the argument, the principle of teleology is assumed to apply, the interaction between the traits of character of the agent with the acts chosen via deliberation would appear roughly as depicted in figure 2:

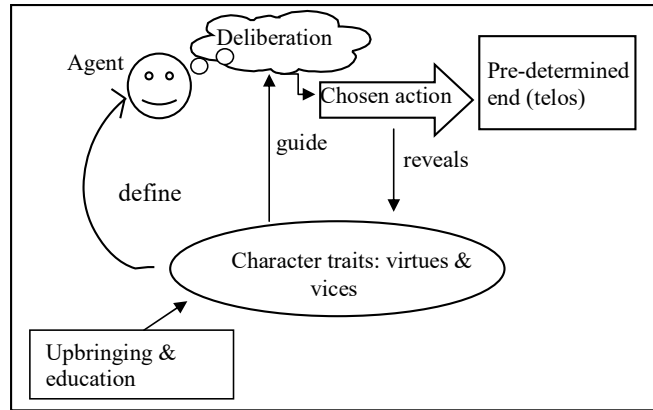


Figure 2 Virtues and the deliberation on action in Aristotle.

In the figure, the deliberation leading to the chosen acts of the actor are guided by his or her character traits, virtues and vices.¹⁸⁸ Prior to this, the effect of the environment through upbringing and education has shaped the catalogue of the actor's virtues and vices. Therefore the indirect effect of the environment is felt in every choice that the actor makes. As Martha Nussbaum noted in *The Fragility of Goodness* (1986): "Aristotle actually writes 'We deliberate not about ends, but about what is *towards* the end'" (Nussbaum 1986, 297). As the chosen act is committed it becomes apparent to anyone observing the act, and thus reveals the character traits of the actor to the public. These can then be evaluated as having had or not had enough influence that the actor himself or herself could be held responsible for the act. As the consequence, dictated as *telos*, in the incompatibilist context would be pre-determined; it would not have an effect on the assessment of responsibility. This concludes the short excursion into the teleological outlook of the general premises of the Aristotelian theory of responsibility for now. Following this comparison, attention is directed back to the naturalist, and compatibilist interpretations of Aristotle and his theory of responsibility. The main difference between the accounts is that instead of being pre-determined, *telos* is determined by the agent's character dispositions in terms of *wanting* (*boulesis*), meaning the process by which the end of the action is selected (Bobzien 2014, 14).

Despite his heavy reliance on the principle of teleology, Aristotle would not have necessarily identified with incompatibilism. Supporting this claim, Susan S. Meyer argues just the opposite in *Aristotle on Moral Responsibility* (2011). Her scholarship suggests that in Aristotle's case the causes can be

¹⁸⁸ Compare this with Bobzien's description of "How deliberation and choice are related" in her article (Bobzien 2014, 11).

traced back to the origin of action, which in his context is the agent. This means that the chain of causality does not extend anywhere “before” the agent, and prevents a claim of universal determinism in his case. (Meyer 2011, 156-158.) As for what follows, the basis of discussing Aristotle’s responsibility here is Meyer’s “roughly naturalist” interpretation of Aristotle’s responsibility. By following Strawson, this makes for an easy comparison, but the same risks of misinterpretation that Meyer notes apply as well as some anachronistic notions (ibid. 161-167).

While action is determined in Aristotle either on the principle of teleology, or at least the premeditated set goals of wanting, development of human character does not follow a similarly straightforward pattern (Roberts 1989, 30). Whether determined or not, i.e., regardless of the principle of teleology, Aristotle’s theory of moral responsibility is incomplete without the notion of character. In talking about character, Aristotle is concerned with effective causes directed toward oneself. In that case, neither nature, teleology, nor wanting have an effect on responsibility. So how can human beings be responsible for their characters? How do acts directed at oneself have an effect on the selection of such acts or those coming after? These questions are discussed in the next section, but a few words will be first given to the concept that was of more particular interest in Strawson, which is *blame*.

As blame is currently an active topic in the philosophy of responsibility, it is useful to look at history of this debate in Aristotle’s case. As with moral responsibility, Aristotle can be considered the originator of theories of blame as well. The concept of blame has a central role in Aristotle’s theory, since right from the beginning of book III of the *Nicomachean Ethics* voluntary action (done not out of coercion or ignorance) is considered to be a candidate for praise and blame (*NE* III 1, 1109b30–1109b33). Blame is thus present in Aristotle’s theory of responsibility from the outset: together with praise, blame signifies the approbation or disapprobation of conduct based on the virtues of the agent. Whether this conduct is judged appropriate or not is decided within the social sphere of the polis,¹⁸⁹ which Aristotle discusses in *NE* as well as the *Politics*¹⁹⁰

In the *Eudemian Ethics* Aristotle describes praise and blame as follows:

Since virtue and vice and the works that are their expressions are praised or blamed as the case may be (for blame and praise are not given on account of things that come about by necessity or chance or nature, but on account of things that we ourselves are responsible for,

¹⁸⁹ At least here, based on Irwin’s remark on how Aristotle intended his theories to have relevance in actual cases where responsibility assessments were needed, Aristotle’s moral responsibility is considered to be applicable in actual social contexts.

¹⁹⁰ The two works are considered to be two halves of a continuous study. This interpretation is based on the work of Simo Knuuttila and Juha Sihvola in the Finnish translations of the works, on which the original research was based.

since if someone else is responsible for something, it is he who gets the blame and praise), it is clear that virtue and vice have to do with matters where the man himself is the responsible source of his actions. (EE 1223a9-16.)

According to Bobzien for Aristotle “virtues *are* the character dispositions that are praiseworthy (NE1103a8-10; cf. 1101b13-15, 30-1). He argues that what makes these dispositions virtues is that they are directed towards the intermediate in action (NE 1106b36-1107a1); and that this is also what makes them praiseworthy (NE 1109b24).” (Bobzien 2014.) The virtues are briefly described in the next section.

In terms of Bernard Williams's “elements of responsibility”,¹⁹¹ Aristotle's praise and blame represent the response element in his theory of moral responsibility. In this sense, essentially, responsibility for action refers to praise and blame as forms of feedback, following the act in question, assuming that the relevant conditions discussed above are met. It is specifically assumed that moral responsibility is conceived according to the natural sentiments, most particularly blame. However, Irwin has doubts about the idea of equating responsibility with just praise and blame. (Irwin 1980 117.) Toward the end of his article, he writes: “To be responsible for an action is not to be the object of those attitudes – sometimes it may be pointless or inappropriate to praise or blame – but being responsible is being a reasonable candidate for these attitudes, the sort of agent doing the sort of action for which praise and blame are normally justified.”¹⁹² (Ibid. 134.) Nevertheless, blame in Aristotle appears as an integral part of his writings on responsibility and its related themes.

Where would Aristotle's blame be located in the classic tripartite distinction¹⁹³ used in Coates & Tognazzini (2014)? Would Aristotle's theory be classified as a cognitive theory? At least the idea that the conduct of the citizen is judged by the virtues would fall into this group. Aristotle's blame could also be interpreted to be conative, given that action in his works is aimed at *telos*-type goals and that the mechanics of actualization or the failure of actualization of such goals could be used as descriptions of conatively directed desires or at least a close analogy.

¹⁹¹ The elements of responsibility were first discussed in chapter 4.3.3.

¹⁹² In this sense, Irwin pre-empts the talk about “standing to blame” that is discussed in chapter 4.2.

¹⁹³ It would seem that the classic tripartite division has provoked some criticism based on recent studies of neurological nature. For example, Klaus Scherer questions the utility of making this distinction by noting among other points that cognition is rarely devoid of affect. Yet the division seems so deep within the Western history of philosophy that picking it apart would be too difficult in this relatively short thesis. At least Aristotle can be straightforward placed within the current selection, as he was taught by Plato, who is originally credited with the tripartite division. (See *Republic*, book 435bd.)

5.1.2 CHARACTER, VIRTUES, AND MORAL RESPONSIBILITY

Instead of direct acts and their consequences, the focus of *responsibility for character* is on the properties of character that the actor possesses, as well as the previous voluntary acts that have produced these properties. In Aristotle's context, these properties of character are related to the effective cause, instead of the principle of teleology.¹⁹⁴

If the beginning of *NE* book III (as well as the *EE*¹⁹⁵) only was consulted, it would be sufficient to summarize the basics of Aristotle's theory of responsibility along the lines of Irwin's simple theory. However, as the critics point out, once Aristotle included book III.5 in the *Nicomachean Ethics*, the big picture became less tidy with the comment that it would be "irrational to suppose that a man who acts unjustly does not wish to be unjust or a man who acts self-indulgently to be self-indulgent".

The full passage, which includes the effect of character within moral responsibility, is:

[F]or it is activities exercised on particular objects that make the corresponding character. This is plain from the case of people training for any contest or action; they practise the activity the whole time. Now not to know that it is from the exercise of activities on particular objects that states of character are produced is the mark of a thoroughly senseless person. Again, it is irrational to suppose that a man who acts unjustly does not wish to be unjust or a man who acts self-indulgently to be self-indulgent. But if without being ignorant a man does the things which will make him unjust, he will be unjust voluntarily. (NE III 5, 1114a 8–14.)

In this observation, Aristotle introduces a new set of considerations into the theory, on which there is considerable divergence of opinion.¹⁹⁶ Throughout the act, then, the unjust or self-indulgent man probably acts as his character dictates. At the moment the act takes place, it is thus possible that the unjust or self-indulgent act is done voluntarily and in full knowledge. Here Aristotle is criticising Plato's view that nobody does wrong knowingly. One can do wrong, and can be held responsible for the act and be blamed for it, even if the actor has full knowledge of the situation (see Roberts 1989, 32–33).

According to Aristotle, the actor is simultaneously responsible for his actions directly and indirectly through his character as well. As per responsibility for action, the feedback of society is aimed at action that is taking place not what the actor has done in the past. In the case of

¹⁹⁴ See G. E. M. Anscombe 1977, 68–69 for comparison.

¹⁹⁵ According to Irwin, considerations of responsibility for character are not included in the *Eudemian Ethics*, but claims related to the topic are made in 1225b11–17 and 1223a5–14 (Irwin 1980, 154).

¹⁹⁶ See Fischer & Ravizza 1998, 12, n17 for comparison.

responsibility for character, the viewpoint shifts to the virtues or excellences of the actor.¹⁹⁷ Taking into account both the actions as well as the moral characteristics of the agent indicates a prospective, forward-looking dimension of responsibility (Sarah Broadie 1991, 124-125). What are these virtues and excellences, and how does the interaction between them and the acts happen in Aristotle's theory?

Aristotle includes virtues (or excellences¹⁹⁸) within his conception of the soul (*NE* II 5, 1105b20–1106a12). Virtues are, depending on the interpretation, a means to the good life or an end in themselves (See *NE* X 6, 1176a30–1176b8). They are states of character that each have their corresponding two vices (*kakia*). One of these means lack of the virtue or excellence and other its excess. For example, courage is the *golden mean* where cowardice is the lack of it, whereas rashness is its excess. Through practice, the virtues or the excellences tend to approach their golden means. In determining responsibility, virtue or excellence appears as a guiding property of character in the Aristotelian account. It decisively affects the selection process of the act to be committed from all possible alternatives leading to the same end. The behavior of the individual is determined by causes of action originating from himself, from his own ethical practice. Nevertheless no one can become virtuous in a vacuum. Could the origin of virtues or excellences be the society? To answer this, Aristotle's distinction between intellectual virtues and those of the character must be considered. Aristotle writes on these different types of virtue:

Excellence, then, being of two kinds, intellectual and moral, intellectual excellence in the main owes both its birth and its growth to teaching (for which reason it requires experience and time), while moral excellence comes about as a result of habit, whence also its name [ēthikē] is one that is formed by a slight variation from the word for 'habit' [ethos]. From this it is also plain that none of the moral excellences arises in us by nature; (NE II 1, 1103a15–21.)

As Roberts's critique (1989) points out, character appears to be a major feature in Aristotle's theory of moral responsibility. Its inclusion is also the main difference between Aristotle's account and the contemporary theories. (See Roberts 1989, 28.)¹⁹⁹

Whereas Irwin focused on the action and the conditions of responsibility, character is present in what Bobzien calls "Aristotle's theory of what makes us responsible for our actions and character" and the reasons for its prominence include setting the deliberation of action within the teleological

¹⁹⁷ Compared to this, responsibility for action was focused on the act leading to a certain end and the praise and blame given by the society based on the result (see 5.1.1).

¹⁹⁸ The words *virtue* and *excellence* are treated here synonymously.

¹⁹⁹ "This is a difficulty for anyone trying to find a hint of a modern notion of moral responsibility here." (Roberts 1989, 28).

context, in which character has sway over the possible choices. (See Bobzien 2014, 2.)

This section provides a broad sketch about how the character is related to moral responsibility in Aristotle's theory. For simplicity's sake, the term "responsibility for character" is referred to as that part of moral responsibility concerned with the questions of character. Character consists of the character virtues as well as their complementary vices for each agent. Aristotle's essentialist metaphysics mean that the same type of virtue that could be shared between individuals was identical in different people. A short summary of what the virtues are is discussed next, and it is then explained how in Aristotle these virtues are connected to the decision-making capacities of human beings. In this case, the relevant concept is *prohairesis*. Finally, some concluding remarks are offered on how the considerations related to character appear in Aristotle's theory of moral responsibility, as discussed in the *Nicomachean Ethics* III.1-5.

Therefore excellence also is in our own power, and so too vice. For where it is in our power to act it is also in our power not to act, and vice versa; so that, if to act, where this is noble, is in our power, not to act, which will be base, will also be in our power, and if not to act, where this is noble, is in our power, to act, which will be base, will also be in our power. (NE 1113b3-7).

In the encyclopedia article on Aristotle (2001), Irwin claims that within Aristotle's theory "[r]ational deliberation [intellectual virtue/practical reason] and decision [*prohairesis*] are the source of responsibility for character no less than for action" (Irwin 2001, 95). In order to frame his argument, Irwin includes two relevant passages. "Virtue and Character" (Irwin 2001, 92-93) and "Voluntary Action and Responsibility" (ibid. 94-95).

The following are true in Aristotle's theory concerning the virtues of character: 1. Virtue is a state (*hexis*), which is a distinct form, but involves both capacities (*dynamis*) and feelings (*pathe*). 2. Virtues are to be followed as the means to good life, but they should also be consciously followed for their own sake in order for the agent's life to be good. 3. The virtues are generally a mean between two vices, but the correct conduct depends on the situation. For example, "extreme anger" can be called for in some circumstances. 4. Moral education is needed in order to align the non-rational part of the soul with practical reason as well as to have better control over the emotions generally. 5. A *prohairesis*-type informed decision and the capability to use it is required for wisdom and the good life. (Irwin 2001, 92-93.)

Myles F. Burnyeat's article "Aristotle on Learning to Be Good" (1999) starts by asking "Can virtue be taught?"²⁰⁰ Burnyeat's article argues that a

²⁰⁰ In Burnyeat's article, "Aristotle on Learning to be Good" (1999) it is asked whether being complete in terms of virtues negates the condition of *akrasia*. According to him this is so. On a related

person who is complete in terms of virtues or is otherwise a perfectly formed moral individual cannot be akratic, as the two conditions are mutually exclusive. (Burnyeat 1999, 224.) In advancing his argument, Burnyeat aptly describes how the practice of the virtues came to form the character of the person. He describes Aristotle's work as a continuation of Plato's pioneering work in moral psychology. Plato and Aristotle agree on one thing, that Socrates was right about education, since moral development should be seen in terms of the conception of virtue. (Burnyeat 1999, 205.) Virtue is thus the key concept within the moral theories of the three Socratic-period philosophers.

Nancy Sherman equates virtues and excellences in her entry on "excellence" in Becker & Becker's *Encyclopedia of Ethics* (2001). She calls for a couple of distinctions to clearly understand the Aristotelian use of the concept. For one, Aristotle's definition of virtue is broader than the usual reference to moral virtues that is used today. Including these, Aristotle's virtue encompasses "any stable state or disposition of a thing which makes that thing do its work well". (Sherman 2001, 504.)

In the entry on "excellence", Sherman also raises concerns over the point in moral development, beyond which the agent moves from non-cognitive habituation to full self-authored agency.²⁰¹ Any definition of such a point is arbitrary and is bound to be controversial. (Sherman 2001, 505.) The point from which childhood changes to adulthood, implying full moral responsibility, is never clear-cut, as Sherman notes that "Aristotle says ethical immaturity can occur at any chronological age" (Sherman 1999a, 237). Following this, it is reasonable in Aristotle's context that young people are supposed to be educated in order to be responsive to rational arguments. Another conjecture from this premise is that rational persuasion cannot have an effect on adults who are "corrupted by a life of pleasure". This could be the way that Aristotle discusses the cases now referred to as psychopaths. (Sherman 1999a, 237 – 238.)

Nancy Sherman's article "The Habituation of Character" (1999a) discusses how the concepts of virtue, moral education and the development of moral agency among children appears in Aristotle's work as well as discussing the relation between virtues, their education and the development of moral agency in their case. Sherman writes about how the development of character and the attainment of full moral status through this process of maturation are fully dependent on the environment in Aristotle's moral theory. This means that character, which is integral to moral responsibility, is social in origin, and therefore no one can become virtuous in a vacuum in his theory. We learn our habits from others and by practice we gain our virtues,

note, thanks to Malin Grahn-Wilder for bringing the intricacies of education of virtues in Aristotle to my attention.

²⁰¹ Dennett's "Plateau" being the obvious example.

as Sherman writes: “Aristotle’s account extends well beyond the truism [...] “We learn by doing””. (Sherman 1999a, 257.)

Children are not to be considered responsible in Aristotle’s context; that much is certain. He does not however underestimate their cognitive capacities.

Susan Meyer’s *Aristotle on Moral Responsibility* describes a feature of the modern theories of responsibility in which the responsibility for character is responsibility for the necessary condition of an action, and goes further in holding responsibility for character the foundation of contemporary moral responsibility in general.²⁰² Voluntary action happens through reflection and deliberation. Deliberation and the chosen act based on it (*prohairesis*) is an essential mechanism of Aristotelian ethics. Virtuous action, as it is understood, is also governed by *prohairesis* choice. (Meyer 1993, 122; see also p. 130.)

The key concept of *prohairesis*, which is based on knowledge and the virtues and which is what children lack, thus distinguishing their moral status from that of adults, is also discussed by Sherman. The child is not responsible because his “desire obeys the reason of the authoritative part as a child listens to his father” (see *NE* 1102b31-1103a3).

Through education, morality can be more persuaded than forced to take hold of the student, and thus it can affect the choices requiring moral deliberation. Sherman discusses Aristotle’s attempts to divide the soul into rational and non-rational. She notes directly afterwards however that Aristotle questions such undertakings in *De Anima* (Sherman 1999a, 235). Sherman is interested in the type of desire that is based on the rational: the rational wish or *boulesis*, which is distinct from appetites and emotions. As in *Eudemian Ethics*, where Aristotle asks “how and by what sources does virtue arise?” he states that “character [*êthos*] is a quality of the non-rational part of the soul (Sherman 1999a, 236). Wanting and wishing is directed at the natural needs and as such are less the subject of choice. *Prohairesis* as the rational counterpart of *boulesis* is more important in terms of moral responsibility. According to Aristotle, the child is incapable of making *prohairesis* choices. (Sherman 1999a, 244).

Summarizing Sherman, Aristotle maintains that children have tools for making voluntary decisions and judgments, and that the task of parents is to persuade the child into a certain understanding about whatever situation he is in, such as behaving in one way is correct in one situation but wrong in another. The point is that the parent cannot make the actual decision for the child. The goal of the educator is not to manipulate, but prepare the child for eventual full moral maturity. In Aristotle’s case, this means developing the means to make *prohairesis* choices. (Sherman 1999a, 242.)

She describes how virtuous actions have their own external ends in Aristotle’s theory, so that in practicing these actions, the practice itself

²⁰² Meyer herself is critical of this idea. (See Meyer 2011.)

becomes virtuous: “For full virtues Aristotle requires not merely that actions be “chosen” [prohairesis] in the above sense, but that they be chosen for their own sakes.” This is possible through the way that virtuous action has ends that are external. It is thus not enough to just choose actions that lead to these ends, but to be fully virtuous one should also act in order to promote those ends as well. Doing so successfully enables the actions to be valued by the external ends in the process.²⁰³ The actions become the ends in themselves. (Sherman 1999a, 244-245.)

Following Richard Sorabji (1980), Sherman describes the learning process of a virtue,²⁰⁴ with first imitating the desired course of action, and through gradual refinement the action approaches its ideal: “in the cases of virtue, the practice of actions will obviously be more complex virtuous action, as we have said, will combine a judgment of circumstances, reactive emotions and some level of decision about how to act. Here too the learner will follow the examples of emulated models and may have in mind general precepts and rules of thumb” (Sherman 1999a, 248). Early learning might be motivated by the pleasures of acting in a certain way and conversely of pain when punished for doing wrong. Later the action becomes its own end, at which point the moral habituation is completed. (Ibid. 256-257.)

Sorabji writes that even though the individual can affect his character virtues by training them, the practice by itself is not enough to make the character virtuous (Sorabji 1980, 211). According to Aristotle, individuals belonging to a species, including humans as representatives of their species, tend to gravitate toward realising the characteristics of their species. Here Aristotle is following Plato’s thought by bringing up the option that the individual could form or acquire habit-virtues through upbringing and/or education, meaning a sort of blueprint for virtue (Sorabji 1980, 211–212). By initially practicing the habit-virtue, the character virtues become determined by individual action and in this way as states of the character they could affect the selection of acts directed at their ends.²⁰⁵ Even if the activity of the actor is needed, the origin of the virtues according to Aristotle lies in the society to which the actor belongs (see Broadie (1991) 170).

According to Aristotle education and the city state are essential in the generation of the intellectual virtues [or excellences]. Aristotle also includes

²⁰³ Here Sherman refers to Eugene Garver’s “Aristotle’s genealogy of Morals” (1984).

²⁰⁴ Nancy Sherman’s “The Habituation of Character” (1999) concentrates on the character virtues: she discusses the nature of virtue, and whether they originate from nature or are taught. Sherman’s article can be read as continuation of some of the themes of Richard Sorabji’s article “Aristotle on the Role of Intellect in Virtue” (1980), as it is mentioned as the second major source cited in her bibliography. Sorabji discussed how the virtues are formed in Aristotle’s theory. Basically, mimicking an outside example to become a habit is a model for behavior that if consistently followed, forming an actual virtue as a result.

²⁰⁵ Seconding this point, Burnyeat notes Aristotle’s Evenus quote: “I say that habit’s but long practice, friend, And this becomes men’s nature in the end.” (Burnyeat 1999, 224.)

practical reason (*phronēsis*) in the group of intellectual virtues. Practical reason is defined by education and experience and will help the individual to achieve the golden mean of his character virtues or excellences. (Sorabji 1980, 210.) By providing education, the society gives the individual the means to attain his ideal state more efficiently. Practical reason thus has an important role in how well the character virtues or excellences are adopted, but it does not directly determine their substance.

Practical reason contains the process of deliberation and reflection, which leads to the selection of the act that fulfils the end. (Broadie 1991, 179). The virtue-guided selection of possible acts is called *prohairesis* by Aristotle.²⁰⁶ The word is difficult to translate, but its meaning is close, as Salkever writes, to a choice made with full awareness of alternative possibilities (Salkever 1990, 69). The concept is so central to the *Nicomachean Ethics* that it is seen as the defining trait of human action (NE VI 2, 1139b5–6). The resulting choice made after deliberation reveals the virtues or vices of the actor to the society around him. Praise and blame are given based on these same virtues and vices.²⁰⁷ It is important to note that since in Aristotle's theory only adult human beings are capable of deliberated, voluntary acts, the definitive threshold of responsibility for character resides at the difference between mature actors and those that are not (e.g., animals and children) (NE III 2, 1111b4–9).

If we accept that the premise that *responsibility for action* and *responsibility for character* describes different aspects of Aristotle's conception of moral responsibility, the resulting understanding can expand the usual picture of Aristotle's responsibility only as the relation between the actor and praise or blame following the act. Responsibility for character implies that the act has counterparts which fulfil the same end. The act, chosen from among alternative possible acts, is selected with the assistance of virtues of character. Because the actor can affect his own virtues with practice, the whole account of responsibility for the act is determined at once by the teleological explanation of the end and the indirect, causal effect(s) of the actor's character virtues.

²⁰⁶ In *Eudemian Ethics* *prohairesis* is introduced book in II.8.

²⁰⁷ In *EE*, this point is made more clearly: "Since virtue and vice and the works that their expressions are praised or blamed as the case may be (for blame and praise are not given on account of things that come about by necessity or chance or nature, but on account of things that we ourselves are responsible for, since if someone else is responsible for something, it is he who gets the blame or praise), it is clear that virtue and vice have to do with matters where the man himself is the responsible source of his actions." (*EE* II.6 1223a10–15.)

5.1.3 THE POLITICAL RESPONSIBILITY OF THE CITIZEN

Even looking solely at Aristotle's writings on moral responsibility, the individual cannot be completely separated from the society surrounding him. The voluntariness of acts and character virtues are not the only factors determining responsibility in Aristotle's theory. In addition to these, it is significant how the action of the citizen relates to the political action of the state. As Delba Winthrop describes it, political responsibility is formed from action, which is defined by civic virtues related to the ends of political action of the state that the citizen inhabits.²⁰⁸ (Winthrop 1975, 417–418.) In moral philosophy, this viewpoint is much less explored than that of moral responsibility and its properties. For example, in political theory the *Politics* is used more often as a source. Ignoring the *Politics* is not necessarily the correct move to make if an encompassing picture of Aristotle's moral theory is wanted. The *Nicomachean Ethics* concludes by remarking that the subject of the *Politics* is to be discussed next. In the *Politics*, Aristotle continues the explication of the requirements for the good life. The aim of political science, the city state, proves to be an essential agent in realising the good life. Both ethics and political science are part of the same whole, "practical sciences". The separation of the two schools was amplified later in the process of editing Aristotle's writings. Before this, the texts of both books were most likely lecture notes of the school of Lykeion.²⁰⁹

Before looking at virtuous action of the citizen, it is appropriate to clarify precisely what state and citizen mean: the state is part of nature in Aristotle's theory. (See Nichols 1992, 15–17.) For this reason, the principle of teleology affects the state as well, and it has its own ends. Virtuous action by an individual citizen directs the political action of the state towards its political end (see Winthrop 1975, 407).

Aristotle defines "citizen" through the state in the *Politics*: "But a state is composite, like any other whole made up of many parts—these are the citizens, who compose it." (*Pol.* III 1, 1274b39–1275a1.) The relationship between the whole and its parts describes Aristotle's essentialist view of the individual.²¹⁰ Even if the traits of the individual cannot be used as such to separate the particular or the person, they are enough to distinguish the person from other forms of life. These traits may be identical between several

²⁰⁸ Delba Winthrop's 1975 article "Aristotle and Political Responsibility" discusses the responsibilities of the citizen and their relationship with the political action of the state, while Mary P. Nichols' *Citizens and Statesmen* (1992) complements Winthrop's view by looking at the various relations between different parts of society and how responsibility functions in these complex social settings. Nichols is a political theorist rather than a philosopher, but her work on Aristotle's *Politics* is useful and excellent, and is used as a summary and a commentary on the most important subjects in that work.

²⁰⁹ See Juha Sihvola's notes of *Politikka* (Aristotle 1991), 221–222.

²¹⁰ See Aristotle, *Metaphysics*, X 10, 1075a11–26 for comparison.

people. The city state was the world the citizen of ancient Greece inhabited, as well as a determining factor which gave meaning to the citizen himself. Citizenship as an essentialist trait was equated by Aristotle with “being a political person”. (See Winthrop 1975, 410 & 413.)

Since good citizenship is implicitly determined by the consistent relation between personal character virtues of the citizen and by his civic virtues, Aristotle thus differentiates the virtues of a good man, which were examined above, and virtues of a good citizen. The former are not necessarily in line with the latter (*Pol.* III 4, 1276b34–35). The virtues of the citizen are determined by how the actor realises action appropriate to his role within the society. Each role has its own set of civic virtues: the ruler has the virtues of a ruler, the soldier the virtues of a soldier, the farmer the virtues of a farmer, etc.

Aristotle can be interpreted to mean that political responsibility is equated with good citizenship. In this case, however, the action of the state is emphasized: the state is like a choir of which the citizen is a member (Winthrop 1975, 418). Based on this, one could assume that action of a good state would require that each citizen act according to his roles as far as possible (see *Pol.* 253).

It is good to keep in mind that the ideas expressed in *Politics* are impossible to apply to the contemporary setting (Nichols 1992, 169). In Aristotle’s society, only the Hellenic class with ownership status had citizenship (see *Pol.* 243). In addition to inequality between the sexes, slavery was a given. Aristotle, in spite of efforts to discuss free men and slaves, and their minds and bodies systematically, did not criticize this situation outright. By this point at the latest, as he happened to extend his incomplete conception of human nature to slaves as well, his view became indefensible (see *Pol.* I 5, 1254a21–23). As a result, he had to discuss *natural* slavery, and so his prejudices against “barbarians” were infused into his political theory (See B. Williams 1993, 110–115).

The virtues of the citizen did not, Aristotle thought, depend on personal virtues. What followed from this was that a good citizen was not necessarily a good person. An exception to this rule was the rulers, who Aristotle thought to have the pre-requisite virtues of a good person, in order to excel in their role (see *NE* I 4, 1095b4–7). The ruler is required to have a developed faculty of practical reason, whereas this ability for deliberation is not essential for an ordinary citizen. Additionally, the ruler is required to understand the subjects’ position, as Aristotle writes “he who has never learned to obey cannot be a good commander” (*Pol.* III 4, 1277b7–16). Only the mastery of his own profession is required of a regular citizen.

Aristotle’s theory states that when acting according to the virtues of the ruler, i.e., when the ruler is in fact good, he advances the political action of the state toward its political end. Political responsibility is realised in his case, because his action is concurrent with the political action of the state. As an interpretation of Aristotle, political responsibility means that the action of

the individual advances the action of the state of which he is a part toward its end.

As in the ideal case of the citizen advancing the end of the state, ideally the state aims to further the well-being of its citizens (*NE* I 7, 1097a15–1098b9 & I 13, 1102a5–10). In this sort of state, the citizens' action as an individual would further the state's objective to serve its inhabitants best. This is how the acts of the citizen related to state's action determine the overall civic political responsibility. As Aristotle writes:

*We maintain that the true forms of government are three, and that the best must be that which is administered by the best, and in which there is one man, or a whole family, or many persons, excelling all the others together in excellence, and both rulers and subjects are fitted, the one to rule, the others to be ruled, in such a manner as to attain the most desirable life. We showed at the commencement of our inquiry that the excellence of the good man is necessarily the same as the excellence of the citizen of the perfect state.²¹¹ (*Politics III* 18, 1288a32–39.)*

Governments include units ruled by one, many or several. Additionally, Aristotle divides these into governments directed towards the end of the well-being of its citizens and those directed at the end of procuring benefits for its rulers. Ideally then, the best political aim of a state is *eudaimonia*, which translates into flourishing or happiness. There can be more than one ruler and ideal governments are called monarchies, aristocracies and polities according to their number, whereas degraded ones are called tyrannies, oligarchies and democracies.

Democracy was seen by Aristotle as the worst government, governed by many, probably because he shared his teacher Plato's opinion. Plato was shocked in his time by the execution of Socrates, a verdict determined by democratic decision. According to Plato, governments degrade from better government to worse in the following order: aristocracy, timocracy, oligarchy, democracy, and tyranny. (Plato, 2007, VIII, 544d–562b & IX, 587c). The division into democracy and polity is made by Aristotle, and is

²¹¹ It is worth considering how this important quote has been translated: the translation by Jonathan Barnes from the 1984 edition is used here: "We showed at the commencement of our inquiry that the excellence of the good man is necessarily the same as the excellence of the citizen of the perfect state." In contrast, Barker and Stalley translate Aristotle in the 1998 Oxford edition as "We have also shown, at the beginning of our inquiry, that the goodness of the good man, and that of the citizen of the best city, must be one and the same." Benjamin Jowett's version in the Random House 1943 edition however reads "We showed at the commencement of our inquiry [n76] that the virtue of the good man is necessarily the same as the virtue of the citizen of the perfect state." Virtue, excellence and goodness have been used throughout the years as synonymous with each other.

determined by the main beneficiary of the government, be it the rulers or the citizens.

How then does responsible action differ between the governments? In good governments, the responsible acts of the individual tend to manifest themselves as politically responsible action as well, meaning that interpretations of responsibility for action, responsibility for character and political responsibility are most closely aligned within monarchies, aristocracies and polities; the benevolent government types.

In states with degraded governments, the situation is more difficult to describe (see Nichols 1992, 87). In these cases, the virtuous action of the individual may very well advance the well-being of its citizens, but from the view-point of the state it can conflict with the ends of the state's political action. In this way it is possible to identify situations through Aristotle's theory wherein actions interpreted by just moral responsibility would seem praiseworthy, but would seem blameworthy when the action of the state is included in the big picture.

Whether the state of the government is tyranny, monarchy or democracy, it affects the whole as virtues affect individuals. Tyranny and other degraded forms of government are analogous with vices and lead to weakened state action in that the acts of the citizens are not directed at their own well-being. This means that the acts of states with degraded governments do not achieve their ideal end, the well-being of its inhabitants, and that an individual acting virtuously otherwise might be considered blameworthy if they were acting against the state's political end. Read this way, Aristotle feels almost modern, and is of great interest as such. As a down-side reading Aristotle thus takes rapid steps toward speculation and beyond the boundaries of our present inquiry. Therefore for now it suffices to point out that it is important to recognize the extended circumstances applying to Aristotle's theory, and that the matter will not be pursued further here.²¹²

²¹² More detail on this can be found in the previous article on the subject by the author (Kaila 2014).

5.2 DISCUSSION: RESPONSIBILITY FOR CHARACTER IN NE III.5 – MORAL RESPONSIBILITY OR MORAL EDUCATION?

Taking stock of Aristotle's theory of moral responsibility, it seems clear that there are multiple interpretations of what is inherent to this concept: the most common, shared understanding of its basics can be summed up in terms of the "simple theory presented in Irwin's article (1980).²¹³ The "simple theory" interpretation, however, leaves out the effects of character discussed in *NE* III.5.

Character, as it was discussed in Aristotle's context, consists of virtues, which have an effect on people's moral choices, which means that Aristotle considers character important for moral responsibility. There is, however, some disagreement on the last point, which is discussed further in this chapter. For example, some authors argue against this (Irwin, Meyer, and Curren) in that by talking about character in *NE* III.5, Aristotle did not mean that past events could decisively alter responsibilities, or that character is not an inseparable component of Aristotle's theory of moral responsibility. but that it is discussed because it is important for moral education (Meyer 2011, 145) and Athenian legal thought (Curren 1989, 259) instead.

One interpretation (see Schoeman 1987, 6 & Pritchard 1991, chapter 2) considered be plausible here, is that Aristotle may not have distinguished between ethics and moral development, and thus saw no problem incorporating the notions of moral education and development into his theory of moral responsibility as well as the more extensive political connotations of the concept. Nevertheless, the important conclusion of the naturalistic interpretation, which is ultimately preferred here, is that blame is the key element of responsibility and as such the dynamics of the social relationships between agents are also relevant in Aristotle's context. (*EE* II.7 1223a9-13.)

Both of Irwin's theories involve an agent doing an act based upon the response that constitutes responsibility. For example, if A drives under the influence of alcohol and injures B, A is blamed for the deed and thus considered responsible. The main difference between the simple and complex interpretations is the added condition of "capable decision". Therefore where in the case of the simple theory the mere voluntariness of consuming alcohol sufficed under the complex theory, for A to be responsible for the act A would additionally require the capacity to decide whether to consume alcohol before driving or not. In most cases, the situations certainly do not differ, but Irwin's complex theory provides additional detail for Aristotle's theory of responsibility for action, which is missing from the common one.

²¹³ See also later publications by Roberts (1989), Broadie (1991), and Meyer (2011).

In his seminal article, Irwin (1980) (1.) disagrees with Aristotle that mere voluntariness is sufficient for responsibility. As stated in his complex theory, he expects the capacity for decisions from the agent as well. Irwin also writes that (2.) Aristotle emphasizes that, once acquired, character is lasting and not easy to change, character is composed not only of beliefs but also emotions as “the result[s] of training nonrational desires”; and noted that: (3.) “Aristotle’s discussion of responsibility for character begins by considering how character is formed by decision and action on it”. Finally, Irwin claims that (4.) his complex theory helps to explain the problems of Aristotle’s topic of responsibility for character or “why Aristotle speaks in these two ways”. He claims that Aristotle seems to think that by deciding the agent can become responsible for more than just the act. (Irwin 1980, 140-141).

The “responsibility for character” aspect of Aristotle’s theory is discussed by Irwin in “Reason and Responsibility in Aristotle” (1980) as well. Irwin interprets the thoughts on responsibility for character present in *NE* III.5 (i.e., “the complex theory”) as not representing a contradiction in terms with the rest of Aristotle’s account of moral responsibility. Irwin demonstrates this by using Hume’s theory, which itself is directly influenced by Aristotle’s theory, as a point of comparison.²¹⁴ According to Irwin, Aristotle’s appended theory succeeds where Hume’s theory fails in that Hume claims that only in-character acts are blameworthy, and all out-of-character acts are excused.²¹⁵ This is averted in Aristotle’s theory as, again following Irwin’s interpretation, a capacity for making reasoned decisions is required as a pre-requisite for being held responsible.

The difference between the two classic thinkers is in how they conceive the *mental states* of the agent, which have an effect on the act and accountability concerning the act. Both of them agree that the condition of causal responsibility has to be true in order for the agent to be a valid recipient of praise and blame. According to Irwin, however, Hume understands this condition in two separate ways: in terms of his or her character and some elements of his person that are durable. (Irwin 1980, 134-135.)

²¹⁴ While Hume’s theory of responsibility is not described in great detail in here, it is used as an example of a theory that uses both concepts of excuse and character in its discussion of blame mitigation. *Blame mitigation* as a term is discussed in the analysis section in chapter 7. Hume’s theory is also an important precedent to Strawson’s naturalism, as Hume shares the outlook of naturalism (see Wilson 2003).

²¹⁵ “It is worth comparing Aristotle’s doctrine with another [...] [of] Hume’s attempt[s]. Hume agrees with Aristotle’s condition (1a) in insisting that someone is open to praise or blame only if he is himself the cause of his actions.” [...] “But Hume understands this requirement in two nonequivalent ways: responsible action must proceed (a) from person’s character and (b) from something durable in him.” (Irwin 1980, 134-135.) See also “animals and children satisfy the second condition but not the first” (Irwin 1980, 135). See also Bayles (1976) for a comprehensive account of Hume’s excuses.

As Aristotle mentions, the reason why children and animals are exempted from responsibility is that *prohairesis* choice is unavailable to them, and thus their characters are considered incomplete (See *NE* III.2 1111b8-9). Hume claims however that the durable qualities of the mind and the character are identical with each other. Irwin claims that Hume is wrong in equating the two, and as a result is incapable of correctly evaluating the situations in which the agent is acting out-of-character. Irwin writes that Hume's theory, compared with Aristotle, fails in cases demonstrated in the article. A person previously known to be honest embezzles some money. Because the act of embezzlement is done out-of-character, this act cannot be used as basis for holding the agent responsible in the context of Hume's theory. In Aristotle's case, using the complex theory the agent would clearly be responsible for the act, as he would have had the capability to decide to act fraudulently in any case. (Irwin 1980, 135.)

Irwin also writes: "The first condition is too restrictive. Hume may be right to say – and Aristotle agrees – that someone will be blamed less for what he does 'hastily and unpremeditatedly' than for what proceeds from his character. But he is wrong to identify every durable element in a person with his character." (Irwin 1980, 135.) Aristotle's deliberation and decision provide a better grounding for judgments than Hume's idea of the durability of the principles of mind does. Aristotle has a larger variety of constant features of the mind, which Hume does not distinguish; he equates every durable element of the mind with character. Irwin gives an example in which the agent clearly does wrong, but in a way that the action does not originate from the character. Hume's theory cannot hold him responsible, while Aristotle's appended with Irwin's "complex theory" can. (Ibid.)

The classic example of *compulsion* is the ship's captain abandoning his cargo in a storm.²¹⁶ The example puts the mental state of the captain under the magnifying glass. The situation is interesting because it assumes that the captain is fully causally responsible in losing the cargo, as well as is acting voluntarily on the level of action. Thus, in terms of blameworthiness the only redeeming factor can come through the impersonal coercion that the environment causes. Resulting from this, the stressed state of mind of the captain can also come into question in how it affects the ultimate decision. According to Irwin, his formulation of Aristotle's "complex theory" can solve a lot of problems associated with the situation. By focusing on the diminished capabilities of the agent, which, assuming that the captain is not able to save the cargo, can lead to the conclusion that the captain is indeed not blameworthy on account of the loss. (See Irwin 1980, 137.)

Irwin concludes that "on this theory [the complex theory] of responsibility, someone can be responsible for an action without being

²¹⁶ Fischer & Ravizza follow up with this example. They describe the situation as an intentional, but non-responsive act. (Fischer & Ravizza 1998, 83.)

responsible for his character or for the traits of personality that cause him to deliberate and act as he does” (ibid. 139).

Children are a group that receives special treatment right from *NE* book III onwards.²¹⁷ Aristotle’s reasoning on why children are discussed relates directly to the question of character in moral responsibility. Whether only adults can be held morally responsible is the probable origin of this clarification.²¹⁸ Irwin discusses Aristotle’s treatment of the moral responsibility of children: “And indeed Aristotle seems to be thinking of adults rather than children when he argues that we are responsible for our characters. [...] Only an adult could be expected to know that his actions will form his states of character.” (Ibid. 1980, 140.)

In endnote 43, Irwin also writes that experience and judgment are required from students of politics, both of which children lack (*NE* 1095a2-11). Those who are lectured on political science must have a good upbringing in Aristotle’s theory (*NE* 1095b4-6). People become good naturally, through practice or by education; however, to become virtuous is unlikely for the impulsive children without guidance, to secure which Aristotle suggest laws (1179b21-34).

In this way, according to Irwin, “Aristotle thinks it is possible to be so badly brought up that he is incapable of becoming good even if he wants to be” (Irwin 1980, 154). However, Sarah Broadie (1991) is against this, arguing that Aristotle is well aware of children and considers cases with undeveloped characters, such as those of children. Just because they lack the capability for deliberation, including *prohairesis*, does not mean that they would not strive to be and do good. Just at the end of the responsibility assessments, blame on children can be withdrawn, but not without consideration of their capacity for moral thought. (Broadie 1991, 169-170.)

Marion Smiley in *Moral Responsibility and the Boundaries of Community* (1992) notes that while action can be committed by adults, children and animals alike, the last two groups are excluded from responsibility on the basis of their lacking status in terms of voluntariness. She criticizes Irwin’s interpretation of Aristotle’s responsibility in that Irwin confuses Aristotle’s concept of responsibility with that of Kant. Voluntariness is something for Aristotle that is mediated by political and social norms, not something that depends on “rational control” or free will. (Smiley 1992, 50.)

²¹⁷ *NE* III.2 1111b8-9. See also Bobzien’s notes about children (Bobzien 2014, 20).

²¹⁸ Sarah Broadie writes in *Ethics with Aristotle* (1991):

“It is sometimes suggested that what motivates *NE* book III.5 is the following reasoning: (1) a person of formed character is *not free* to act otherwise than in accordance with it; (2) one cannot be held answerable for doing X when one is not free not to, except on the condition that (3) one freely and knowingly entered into the situation in which one is not free not to do it. With this in mind, it is suggested, Aristotle makes the acquisition of character ‘depend on us.’ For then the fully formed agent cannot escape responsibility for his actions.” (Broadie 1991, 170.)

Aristotle's advantage, Smiley continues, comes from how the practice of blaming children and animals (as well as the mentally impaired) is independent of the status of the voluntariness of the actor. If voluntariness is held as an absolute prerequisite for blame, we lose the details of different levels of voluntariness, and how it is related to the ultimate sanctioning. According to Smiley, Nussbaum's description of the "mysterious shift" between childhood and maturity is something Aristotle's theory is better equipped to deal with precisely because of the separateness of voluntariness and blameworthiness (Smiley 1992, 51.)

Susan S. Meyer considers Aristotle's theory of responsibility and its relationship with the concept of character in her book *Aristotle on Moral Responsibility* (2011). Meyer, whose conception of responsibility is adopted from Strawson, writes that Aristotle's responsibility for character is not a part of his view of moral responsibility, but of his ideas of moral education. She does however note that the modern writers often consider Aristotle's responsibility for character essential to moral responsibility. (Meyer 2011, 145.)

The initial idea of *Aristotle on Moral Responsibility* (Meyer 2011) is to describe the part of Aristotle's work that corresponds to the Strawsonian conception of moral responsibility and argue for a coherent Aristotelian concept of moral responsibility. Meyer (2011) is thus optimal in identifying the similarities and the differences between the different conceptions. She relies on the passages devoted to exploring the notions of voluntariness and responsibility for character. Her inquiry goes to considerable lengths in examining the relationship between voluntariness and responsibility, even so far as responsibility not being a direct concern of moral agency. Her main target is Roberts' article (1989), as she objects to Roberts's stance as one where the voluntariness aspect gets neglected, while the particular effects of character are agreed upon (see Meyer 2011, 14).

Meyer describes accounts of moral responsibility in terms of the scope of their definition: moral responsibility in Meyer's narrow construal is the "concern of criminal law [...] It concerns the conditions in which an agent merits praise and blame, or reward or punishment, for an action. An agent merits blame or punishment for an action if and only if she is morally responsible for it [...] On this narrow construal, moral responsibility is a relation between an agent and her action". (Meyer 2011, 17-18.) The narrow construal thus concerns only the retributive side of the concept. Compared to this, Meyer, presents moral responsibility at its broadest construal as "issues of central importance to our conception of morality and to our conception of ourselves as moral agents". (Ibid. 18.) Meyer's special interest is aimed at "agents who are subject to the demands of morality", while she makes the usual exceptions of "non-human animals, small children, and the mentally defective". She states that while concern about these groups is important, "we don't consider these claims to be reciprocal", which corresponds to

Strawson's distinction between the objective attitudes and reactive attitudes (ibid. 18).

Meyer writes about these statements that recall Aristotle's emphases on virtues and rationality. They fit Eshleman's categories in chapter 3, and her work can be said to be influenced by the Strawsonian tradition.²¹⁹ Going through the criteria reveals an idea of responsibility, at least in stating the scale between the narrow and broad construal (ibid. 17-18): there is an agent²²⁰ (ibid. 28), there are conditions, which determine whether the agent can be held responsible for an action in terms of the exceptions on p. 18, as well as the statement that voluntariness not a sufficient condition of responsibility (ibid. 28-29). Finally, there are the virtues, which act as objects of responsibility ascriptions (ibid. 19-24).

The focus on virtues lets Meyer concentrate on responsibility for character.²²¹ Throughout the work she is interested in the relationship between the notions of responsibility for action and character.²²² She makes the effort to emphasize the role of moral education as a separate but related area with moral responsibility in general. She takes the risk of attempting to include the former within the latter clear, and is skeptical about such projects. She writes in the introduction:

Aristotle clearly has the concerns of the moral educator in mind when writing his ethical treatises and he thinks praising and blaming play an important formative role in moral education. But he nonetheless conceives the goal of moral education to be to produce a fully autonomous individual who merits praise or blame. The only agents who merit praise or blame are those who are properly subject to the demands, expectations, and evaluations of morality. The first task of a theory of moral responsibility is to identify the features that properly subject an agent to these demands, expectations and evaluations. An agent who has these features is a morally responsible agent. Aristotle's account of moral character (the condition common to virtue and vice of character) accomplishes this first task of a theory of moral responsibility. (Meyer 2011, 3-4.)

²¹⁹ Meyer refers directly to *FR* (Meyer 2011, 17-19, endnote 1 & 2.)

²²⁰ Meyer discusses Aristotle's moral agency in the sixth chapter of her book: "the agent is not simply a causal factor in production of the action, but the causal factor to which primary causal responsibility is attributable" (Meyer 2011, 149).

²²¹ For example, Jean Roberts' article (1989) is cited in the introduction (Meyer 2011, 2, endnotes 4 & 5.)

²²² E.g., "The state of character concerning a given range of activity disposes the agent to display or engage in that activity in a certain way in certain circumstances. The justice or injustice (e.g., flight or fear) depends essentially on these additional conditions "What, when, how, etc." in Aristotle's formula [...] Hence the thing for which one is blamed or praised is not, for example, getting angry, but rather getting angry in these circumstances to this degree, etc" (Meyer 2011, 163, *NE* references excluded).

Meyer continues that Aristotle's theory also leaves room for the exceptional cases, as while he "explicitly recognizes that not all voluntary agents are morally responsible for their voluntary actions, he nonetheless thinks that agents with moral character are morally responsible for their voluntary actions". (Meyer 2011, 4.)

The emphasis on the voluntary is not uncritical in Aristotle according to Meyer. There are examples where Aristotle considers alternatives to the most commonly used schemas.²²³ Meyer applies Strawson's view on compatibilism to Aristotle too, as she states that the question of determinism is not a concern nor is definitive with the topic of moral responsibility (ibid. 4). According to Meyer, moral character means that a "set of dispositions together determine the full range of agent's sentiments and interests" (ibid. 31). With a reference to Roberts (1989), she describes responsibility for character as a *prospective* type of responsibility (Meyer 2011, 40-41).

Aristotle, according to Meyer, "thinks the qualified sort of responsibility for character [...] we have is significant for our moral education, not for our moral responsibility." (Ibid. 145.) This is coherent with her main argument that Aristotle's responsibility cannot be interpreted solely on the basis of voluntariness. Responsibility for character might not be required for responsibility for action, as Meyer notes: "Aristotle nowhere indicates that he thinks responsibility for character is significant because it is a necessary condition for responsibility for action, or because it is necessary for the praiseworthiness or blameworthiness of states of characters [...] in this context [arguing against the Socratic asymmetry thesis] he thinks the qualified sort of responsibility for character [...] we have is significant for our moral education, not for our moral responsibility (n28)" (Meyer 2011, 145. See also Roberts 1989, 30-31 and Broadie 1991, 166-74.)

In an endnote Meyer compares Robert's and Broadie's position to her own: "Roberts [1989, 30-31], like Broadie [1991, pp. 166-74], agrees that the point of Aristotle's argument for responsibility for character is to establish that nature leaves room for habituation and education. But she takes this to indicate that Aristotle is not concerned with moral responsibility in the chapters in which he discusses voluntariness" (ibid. 148, n28). She interprets Robert's position as agreeing with that of Bernard Williams's *Ethics and Limits of Philosophy* (1985). (Meyer 2011, 148, n28 & n29.)

Following a quote from *NE* (1114a4 - 13), Meyer situates the character's effect on moral responsibility most clearly thus: "Modern accounts of moral responsibility often take responsibility for character to provide the ultimate basis of moral responsibility." (Meyer, 2011, 122) She cites Williams 1985, Roberts 1989 and Broadie 1991 as sources for this statement.

Meyer likens *moral character* to the immovable mover or voluntary actions of the moral agent within Aristotle's theory (Meyer 2011, 154-155). Her interpretation of Aristotle's moral agency "may be roughly described as

²²³ One such example is the diachronic model of action mentioned briefly in the *Metaphysics*.

'naturalism"', as natural processes are not up to us (ibid. 160, see also 188). She notes that the current theories would accept the idea of *agency* as an alternative to explanations related to character, i.e., the causal features of moral agency would be exceptions or interruptions (ibid. 161).²²⁴ Since character is the non-accidental cause of a desire, and as non-accidental causes exist in nature in Aristotle, this is not sufficient for determining a responsibility relation (Meyer 2011, 160).²²⁵

Randall R. Curren, whose 1989 article "The Contribution of *Nicomachean Ethics* iii 5 to Aristotle's Theory of Responsibility" discusses the influence that the titular chapter has had on interpretations of Aristotle's theory of responsibility.²²⁶ While Curren claims that Aristotle's conception of responsibility is not compatible with the modern counterparts including the Strawsonian conception (see Curren 1989, 261-262), he demonstrates that chapter III.5 has a significant effect on moral philosophy and jurisprudence that has appeared since. (See Curren 1989, 274-275.) Included in this evaluation are the specifics of character and their relation with responsibility.

On responsibility for character, Curren writes: "So for Aristotle an agent may be responsible for something without having decided to bring it about or having foreseen its coming about, provided its source and cause is the agent's character acting through the mediation of the agent's intellect." Ultimately Curren rejects the attempts to include concerns of character within the scope of moral responsibility (Curren 1989, 266). His article has some interesting conclusions as a result. According to Aristotle, even if we cannot choose our ends, III.5 implies that we can "contribute to" them through voluntary action. (Curren 1989, 274-275.)

The way that Curren ties blameworthiness to voluntariness as Irwin does is equally subject to Smiley's criticism. A bigger problem for Curren's article itself is unfortunately that it is compromised by the heavy reliance on *Magna Moralia*, the authenticity of which has been questioned. The reason why

²²⁴ As is the case with the Strawsonian theories examined in the chapters above.

²²⁵ According to Meyer, Aristotle holds that "only causal claims worth taking seriously concern non-accidental causal relations, and explicitly includes his discussion of voluntariness within the scope of this claim" (*EE* 1221b4-6, in Meyer 2011, 157). "While rocks, floods, viruses, non-human animals, children, and adult humans can be causally responsible for results, we suppose that only adult humans are morally responsible for the results they cause." Meyer continues: "While [Aristotle] clearly thinks that entities other than morally responsible agents can be non-accidental causes of outcomes, he thinks that only certain entities can be non-accidentally the cause of morally significant outcomes." (Meyer 2011, 162.)

²²⁶ Randall R. Curren's article "The Contribution of *Nicomachean Ethics* iii 5 to Aristotle's Theory of Responsibility," published in *History of Philosophy Quarterly* (1989) is included as it specifically discusses the meaning of chapter iii.5 of the *Nicomachean Ethics* to Aristotle's theory of moral responsibility. Curren is among the skeptical writers who would like to keep the themes of *NE* III.5 separate from Aristotle's theory of moral responsibility, as he considers the chapter is related to Athenian legal thought instead.

Magna Moralia and the *Eudemian Ethics* to lesser extent are emphasized is the hope of showing that Aristotle's theory of responsibility is not solely reliant on the concept of voluntariness as defined in *Nicomachean Ethics*, as the discussion is centered around involuntariness in that context (see Curren 1989, 262).

Burnyeat sums up the thoughts on moral education in *Nicomachean Ethics* clearly. He notes that identifying the workings of moral reasoning alone as moral philosophy is not sufficient. What is often left out of these theories, including those about responsibility, is the subject of the moral development of a person.²²⁷ One of the questions that Burnyeat aims to answer is whether a person who has the optimal set of virtues can be akratic²²⁸ at the same time. Aristotle's picture of moral development is hardly perfect – as Burnyeat acknowledges – but it is ground-breaking and rarely superseded as an account of its kind (Burnyeat 1999, 221).

Further, according to Burnyeat: "What is exemplary in Aristotle is his grasp of the truth that morality comes in a sequence of stages with both cognitive and emotional dimensions." (Burnyeat 1999, 221). For example, the rapid responses of the spirit that are described in *NE* VII.6 (1149a26-b3) contain the concept of "evaluative responses", which are close to the reactive sentiments described by Strawson. Michael S. Pritchard is of the same mind as he describes Aristotle's view on the developing responsibility in chapter 2 of his book *On Becoming Responsible* (1991).

Two interpretations of Aristotle's moral responsibility

Aristotle's concept of moral responsibility can be found in *NE* book III, and can be interpreted as part of his moral theory that encompasses the nature of individuals and their psychology and politics as well. While a complete reading of all the relevant works (*NE*, *EE*, *Politics*, *Rhetoric*) would present all the considerations needed to represent the wealth of Aristotle's effort in its entirety, the specific concept of moral responsibility seems to have meaning even when discussed independently of the extended notions of political responsibility. The majority of writers are in agreement on this. The critical difference of opinion is related to the effect of character on moral responsibility. At least two major options of interpretation are being debated among the scholars:

²²⁷ The actual theories of developmental psychology have not had a lasting impact on moral philosophy at least during the 20th century. Burnyeat's assessment is close to those of Fingarette (1966) and Pritchard (1991) in this matter.

²²⁸ Being akratic or weak-of-will means that the person fails to achieve his ultimate goal because of his nature (*telos*), even when the actualization of it is something that he consciously wants.

Option 1.

Aristotle's moral responsibility consists of:

- a) responsibility for action (A does X voluntarily (in knowledge, uncoerced, is the recipient of blame, etc.)
- and b) responsibility for character (partially self-trained virtues and vices direct *prohairesis*)
- thus A's blameworthiness is determined by both a) and b).

Option 2.

Aristotle's moral responsibility consists of:

- a) responsibility for action (A does X voluntarily (in knowledge, uncoerced, is the recipient of blame, etc.) only!
- Therefore while questions of character (partially self-trained virtues and vices directing *prohairesis*) are important before the agent is a fully capable adult human being, they do not affect the blameworthiness of A's action directly or indirectly.

The psychologically minded Schoeman and Pritchard make claims that would adhere to the statements in option 1. At least Irwin, Meyer and Curren would go for option 2. The contrasting example of Hume's responsibility (discussed by P. Foot 1966, Irwin 1980 and P. Russell 1995, chapter 7), which takes option 1 literally, would be at odds with option 2. If considerations of character are to be taken into account, they cannot be definite, which is worrisome for proponents of the first option, since the uncertain nature of character is built into the concept of moral responsibility as well in that case.

If we accept Meyer's claim that responsibility for character is *not* significant for moral responsibility, and we opted for option 2, then how would it affect the special cases mentioned in the examples? If the agent is a child, for example, should not the relevant character-based concerns be a part of moral responsibility, because without this it would be more difficult to determine that she is not evaluated under the same terms? Conversely, should not the basic qualifications for the status of moral agency be evident within the rules of the theory? Since in Aristotle's case these qualifications are presented in terms of character, could it be said that these are *only* relevant to the topic of moral education and not that of moral responsibility? This all depends of course on the definition and scope of the concept of moral responsibility one has.

The stance taken here follows Meyer in the sense of how she discusses Aristotelian responsibility clearly in terms of responsibility for action and responsibility for character, but on that whether she is right about the exclusion of responsibility for character from Aristotle's big picture of moral responsibility, her conclusions about this matter are not followed. Smiley's recommendation to view blameworthiness as distinct from the prerequisite

of voluntariness is considered worthwhile. Going back to the Strawsonian context, if it is indeed impossible to hold any *deep* notions of moral responsibility as Bernard Williams claimed, it would certainly seem so. However, with the inclusion of the ideas of historical responsibility the issue becomes complicated once again, as Gary Watson showed. In any case, further investigation between the concepts of moral responsibility and character is warranted. This investigation is undertaken in the latter part of this dissertation.

5.3 SUMMARY

The idea of moral responsibility found in Aristotle seems to include two viewpoints. The first, responsibility for action, focuses on voluntary acts (done in knowledge and uncoerced) and on the response to these, or praise and blame. The second, responsibility for character, is based on choices influenced by the character of the individual as well as the states of character defining the individual, which in turn is defined by voluntary acts preceding the morally evaluated act in question. Whether the second viewpoint is accepted has been debated. No matter what the conclusion is, these two interpretations show that the big picture of Aristotle's moral responsibility is something more complex than just causal responsibility for the act and its consequences.

The idea of political responsibility, which can be found separately in Aristotle's works, implies that the compound action of its citizens forms the political action of the state. The ideal end of this action is the well-being of its citizens. If the state is optimally organized, the action committed by the citizens produces their own well-being. When considered together, moral and political responsibility imply interaction between the state and its citizens, both the ability to work together to attain happiness for oneself and one's fellow citizens through the instrument of the state, and the inability to do so.

Aristotle's moral and political theory thus provides an example of something that modern theories that are more inclined to liberalism and individualism do not possess:²²⁹ responsibility cannot be assigned just by looking at the actor himself, but necessarily some responsibility redounds in all cases to the society to which the individual belongs. In Aristotle's case, the political is present in his more pragmatic version of the theory in a very real way, but as a way of doing ethics the comparison between the theories and the reality quickly becomes an impossible task. For this reason, neither the line of thought involving actual political implications, nor the precise concept of political responsibility are discussed beyond this point, except of course for the ways that responsibility is generally relevant to the social nature of humanity.

²²⁹ Compare with this with Smiley 1992, 34-36 and Nichols 1992, 1-12.

End of Part I “What does responsibility mean here?”

Now that the first research question of (Q1) “How is responsibility understood by philosophers today?” has been answered, first, by introducing the concept through the popular article “Freedom and Resentment” (2008/[1962]) by Peter F. Strawson, and discussing the theoretical framework that has been formed under its influence, and second, by describing its Aristotelian counterpart and the relevant issues related to it, a brief summary of the first half of the dissertation is in order.

By no means has it been claimed that a complete account of theories of moral responsibility or indeed that a straight-forward definition of the concept itself has been constructed. The theory of moral responsibility based on Strawson’s article alone has left a lot of questions unanswered. The idea that his relatively brief text on the subject has resulted in a complex theory that could address every situation can be set aside. Nevertheless the debate provoked by this particular, influential attempt to end one of longest arguments in the history of moral philosophy (i.e., the problem of determinism) is sufficient in that we can at least speak of “a position”.

To flesh out this position further, the foremost current Strawsonian theorist of responsibility has been discussed: R. Jay Wallace’s *Responsibility and the Moral Sentiments* (1994) does a reasonably good job of putting all the pieces of the Strawsonian puzzle together, but he does (inevitably perhaps) introduce new flaws into this position.²³⁰ The problem that was thought to be the most difficult was the conflict between demands that Wallace dealt with regarding questions of responsibility without a separate, metaphysical level (a trait which Strawson introduced to the debate on responsibility) and his (Kantian-flavored) reliance on the particular “morality system”, which Paul Russell pointed out in his critique (2013). For example, we asked whether or not the excuses as “blameworthiness inhibitors” were not part of a separate layer of (cognitive) knowledge that would supervene the reactive sentiments.

Next, the challenges to the Strawsonian concept of blame, including those mounted by Hieronymi, Scanlon and Sher, were discussed. It was noted that the academic debate on blame would continue, and that it would certainly have an effect on future of moral responsibility as well. As for Strawsonian theories in general, it was finally settled that blame is a significant component of responsibility and that the composite moral responsibility in Strawsonian terms is thus dependent on the reactive attitudes (including moral sentiments such as resentment, indignation and guilt) that the members of the moral community display toward each other.

As the criticisms of Strawson and his followers have claimed, there were weak spots in his sketch of the concept. Developmental psychology dealing

²³⁰ At this point, we note that an alternative interpretation of the Strawsonian position on moral responsibility could be made.

with moral development and other factors leading to situations in which the agent is considered extraordinary by some measure were seen to be insufficiently approachable with the Strawsonian mind-set. Wallace's 1994 exemptions, named after Watson's important contribution (1987) are the apparent blind spots of the theory. When the mutually reactive "webs of human attitudes and feelings" collapse or cease to exist, whenever such extraordinary agents are discussed, it is clear that a more comprehensive account of the subject is called for. Furthermore, the general tendency to distance itself from "how we actually treat each other" troubles the Strawsonian position as well. Smiley and others criticized the Strawsonian theory for giving up too much of its social relevance.

As a result of the heuristic comparison, based on *FR* and the texts of Watson (1987) and Wallace (1994), it can be said that the Strawsonian theories in general:

- a) view reactive attitudes and sentiments as constitutive of moral responsibility,
- b) consider affective blame important as a distinctive sign of holding someone responsible and
- c) accept excuses in certain situations as over-riding elements of responsibility resulting in modified or eliminated blame.

Aristotle's theory of moral responsibility is a valuable counterpoint to the Strawsonian position. It was interpreted here from the same broadly naturalistic perspective as the Strawsonian theories were and, by making this comparison, it was acknowledged that the Strawsonian approach to the concept influences the current academic discussion on Aristotle's responsibility. Susan S. Meyer's book *Aristotle on Moral Responsibility* (2011) in particular was referenced as it examines the Aristotelian content of the topic from a Strawsonian vantage-point. Articles by Terence Irwin (1980) and Jean Roberts (1989), Sarah Broadie's book (1991), Garrath Williams's bibliography (2010) and most recently Bobzien's article (2014) reinforced the basic idea that Aristotle's theory of moral responsibility is situated primarily within book III of *Nicomachean Ethics*, and needs to be read specifically with the knowledge that chapter III.5 contains Aristotle's account of character's relation to moral responsibility.

The backbone of Aristotle's theory of moral responsibility is formed by his theory of responsibility for action (from NE III.1 onwards). The most in-depth study of this subject can be found in T. H. Irwin's "Reason and Responsibility in Aristotle" (1980). Irwin summed up Aristotle's theory as "A is responsible (a candidate for praise and blame) if and only if X is done in knowledge and without coercion" (Irwin 1980, 125). Without going deeper into Irwin's additional interpretation (that the "complex theory" of Aristotle's responsibility requires the capacity to decide rationally on the act) it can be accepted as sufficient.

The character-element has, however, proved to be controversial. In what way did Aristotle mean that the agent is responsible for his character as well

as his actions in *NE* III.5? Some authors, including Irwin and Meyer, objected to the inclusion of the notion of responsibility for character in Aristotle's theory. Meyer called these character-related notions in Aristotle more important to his theory of moral education than that of moral responsibility. Others such as Roberts (1989) and Curren (1989) questioned the entire theory of responsibility or at least as it is understood today, based on the imprecise concept.

Aristotle himself seemed to think that by influencing one's own character, the agent could be responsible prospectively for the act, or alternately be excused from blame. On the other hand, a person whose character is known in advance could not be held responsible if he was expected to act otherwise. As for a similar heuristic conclusion on Aristotle, as was done with Strawson above, most commentators on Aristotle's theory of responsibility think the following are true:

- a) In terms of sources, the core of Aristotle's theory of responsibility is located in *Nicomachean Ethics* book III.1-5;
- b) it is based on Aristotle's theory of action and voluntariness;
- c) it contains or is at least related to the notes on character in book III.5,
- d) which describe the reasons for exempting actors from responsibility within its context.

Incongruity between the ways that character is dealt with either in classical or modern theories of responsibility provoked the formulation of research question Q2 "How should character be understood in the context of the philosophy of responsibility?" As neither Strawson nor Aristotle provide a clear and distinctly articulated definition of responsibility, the scholars and researchers working on this have had to debate the subject constantly. The "responsibility for action only" side has, possibly for practical reasons, remained the most usual way to talk about the concept but, despite this, few theorists would wish to limit our view of responsibility to the "ledger-view" that would result from referring solely to this aspect of the term.

As a sort of aporetic conclusion to the dissertation process aiming to answer Q1, we turn to the contribution to the subject matter by Bernard Williams (1993 & 1997) and his sentiment that no single theory of responsibility could be formulated that could satisfy the expectations that philosophers have on the subject. Williams does not however abandon it as something impossible to conceptualize. Instead he suggests four elements of responsibility (cause, intention, state and response). Regarding the previously noted difficulty of incorporating the notion of character into the modern interpretations of the Aristotelian theory, or alternatively as the "blind spot" in the Strawsonian counterpart, these complex subject matters are relatively easily classified within Williams's "state" element of responsibility. This element serves as a crucial point of comparison in chapter 7, but before this explication is properly conducted, this study deviates slightly in the next part. Instead of the historical presentation of the relevant works and the criticism examined so far, a more systematic

exposition of the concepts of character and excuses will make the final analysis of these concepts in the theories of the philosophers discussed above more appealing.

PART II: CHARACTER AND EXCUSES IN THEORIES OF RESPONSIBILITY

The main difference between the Strawsonian and Aristotelian conceptions of responsibility that was noted in chapter 5 relates to the role of character in influencing responsibility. While Aristotle's writings on responsibility seemingly include considerations on the effect of character on the praise- and blameworthiness of the actions, Strawson's article is not interested in character, only mentioning it briefly, and dealing with the special considerations instead. In this chapter, closer attention is paid to the key concepts that are first discussed in terms of Aristotle's responsibility for character. It is demonstrated that the excuses deal with essentially the same element, in responsibility theories.²³¹

In the previous century, the attention of moral responsibility theorists shifted away from character-based accounts as the mainstream of analytic ethics in the English-speaking areas strove for more accurate representation in logic and language (after the logical positivists and Wittgenstein). Recently, however, this tendency has seen another shift arising from the re-surfacing of the virtue ethics of G. E. M. Anscombe (1958) and A. MacIntyre (1981), among others. The sentimentalist work of Strawson, which initiated a development that positively affected a "naturalistic turn" in ethics after the turn of the millennium (with the help of advances in empirical sciences), has been enough to re-ignite an interest in the area, which has been discussed under the topic of "character". There is no clear, unanimous understanding of the topic, which is thus an active field of study. Here the concept of character is first discussed using Marcia Homiak's *Stanford Encyclopedia of Philosophy* article on the subject (Homiak 2015).²³² The contrasting concept of *excuse* is examined because of its similar function mainly among the Strawsonian theorists on blame. The excuses are used to point out exceptional cases of responsibility assessment, where the agent can be found extraordinary because of some fact about her moral status or some external circumstance. For example, the agent could be a child, or her reasoning could be incapacitated due to an acute illness, and so on.

As observed in chapter 5.1.2, responsibility for character implies that the act has alternatives that fulfil the same end and that the agent has a choice. The act, which is chosen among the alternative, possible acts, is selected with the assistance of character virtues. The ability of the actor to affect his own virtues through practice, responsibility interpreted in this *thicker* way, entails that the act is determined simultaneously by the end result and the

²³¹ The word "element" mentioned here refers to Bernard Williams's state element of moral responsibility, as discussed in chapter 4.3.3.

²³² Marcia Homiak's *Stanford Encyclopedia of Philosophy* article on the topic of "Moral Character" (2015) is a great source for the current discussion on character. In addition to a historical overview, Homiak takes into account the critical points that modern psychology has offered against this concept. She examines some of the points made by the critics (most particularly the proponents of psychological situationism) and argues against their points in order.

indirect, causal effects through the actor's character virtues. Thus there is a parallel between the character-inclusive account of Aristotle's responsibility and the common incompatibilist arguments. This question of whether the example of historical responsibility invariably leads to an incompatibilist account is questioned here, as it was first in Watson (1987) as he doubted that it was the case in his Harris example. As for Aristotle, Susan S. Meyer argues that Aristotle's account is in fact compatible with modern compatibilist theories of responsibility. (See Meyer 2011, xxi-xii.)

Having now concluded the two large substance sections of part I, part II will concentrate on the systematic analysis of character and excuses in the context of the theories of moral responsibility. The following two sections (6.1 and 6.2) are devoted to describing the main concepts of *character* and *excuse* and they are analyzed together in chapter 7 in the context of Aristotelian and Strawsonian theories. Where the previous chapters focused on certain theorists and their particular works, the examination that follows concentrates on the specific concepts of interest. This systematic exposition of these concepts in the next chapter (6) advances such that a variety of stances on a concept is first surveyed and, immediately following a critical discussion, a definition is selected for that concept. In the case of "character", Marcia Homiak's *Stanford Encyclopedia* article on the subject is used as a reference, whereas "excuse" will be dealt with as a topic receiving refinement throughout the latter half of the previous century. Starting from Austin's article (1956/57), the discussion advances to Strawson ([1962]) and finally his followers. Once the discussion about the concepts has been concluded, we will examine two examples throughout this work that highlight the phenomenon of blame mitigation.

Chapter seven will contain the analysis, which will contrast the key concepts already presented, and answer the research questions. By utilizing the distinctions of elements of responsibility made by Bernard Williams, a common framework for the two historically separated types of theory of responsibility is made available. Because this systematic assessment of the concepts arises from the sources examined in part I, there is bound to be some overlap. While unnecessary repetition is avoided, some reminders are offered since the matter related to these concepts is considered to be essential.

Chapter eight will contain the concluding remarks of this study, including a summary of the main argument by stating the thesis in compact form (8.1), as well as a broader critical discussion on the overarching topic (8.2).

6 CONCEPTS OF CHARACTER AND EXCUSE

Starting with Aristotelian theory, the means of blame mitigation are conceived in terms of character. Compared to the role of emotions or passions, the states of character enjoy a more stable status within the psychological realm as now understood. Virtues and vices as guiding principles of the mind have a decisive effect on the selection of the action (in the *Nicomachean Ethics* and the work it has inspired) and thus contribute to the result for which the agent is potentially blamed. Whether this trait was self-caused; for example, gained through training, or whether the upbringing of the agent brought its formation about, all this helps to determine the deeper aspects of the determination of responsibility of the agent (including blameworthiness).

Whether this has an effect on the act of condemnation is related to the character in that an incomplete character, lacking the necessary maturity may be grounds for pardoning a wrong-doing. In more complex scenarios, for example, an agent, publicly known for being vicious, can be appointed an obligation to do something important. When meeting this obligation predictably fails, the blameworthiness of the agent can be questioned, and those responsible for selecting the agent for the task could fall under suspicion.

- Aristotle's responsibility for character

Once Aristotle's responsibility for character and its components are sufficiently dealt with, the Strawsonian excuses are compared to them. The similarities and the differences, as well as the arguments that speak for and against the use of excuses instead of *deeper* constructs such as the character, are discussed next.

- Doris's Rejection of Character and the consequent debate

As the broadly analytic Anglophone ethics of the 21st century have embraced the naturalistic turn and are fiercely debating the effect of empirical evidence on certain philosophical issues, such as blame, John Doris takes a multidisciplinary approach in his *Lack of Character* (2002).²³³ Doris rejects

²³³ John Doris's book *Lack of Character* (2002) is a very important book on the discussion about the concept of moral character. It is used here as an example of criticism of the use of character in responsibility assessments. Even Doris does ultimately accept character in a way that allows its

almost all uses for character in ethical discussion related to moral responsibility. (Doris 2002, 1; see also 131.) Characterological accounts according to him serve for little but to clump together complex issues and confuse further. Relying on social psychological studies, Doris argues that moral responsibility cases are invariably situational by nature, and cannot have external structures of morality (including the idea of character) informing them.

Despite this, Doris does allow for an everyday-language use for the concept of character (Doris 2002, 131), arguing that it is natural to use the word to describe the sort of immediate assessments of personalities of others when we interact with each other. This allowance leads to further questions about whether one could append the theory with such notions as Sher's or Scanlon's ideas on social interaction. These speculations will be kept brief, however, as the analysis of the main problems will form the substance of chapter 7.

- Four Definitions of Excuses

Austin started the recent trend of calling excuses as a tool of moral philosophy. His article "A Plea for Excuses" (1956/57) constitutes the starting-point of the debate here. Peter Strawson's "Freedom and Resentment" (2008/[1962]) can be read as taking part in this debate by discussing pleas or special considerations (to the reactive sentiments) in, wherein he listed the conditions that could affect the responses. (See chapter 3.1.3. and Strawson 2008, 7-10.) His notes on the inhibiting effect of traits originally related to agency have since been discussed by Watson (1987) and Wallace (1994), who have called the moderate and total modifications of reactive sentiments excuses and exemptions respectively.

Having contrasted the two alternatives of mitigating blame in the theories of responsibility, Holly M Smith's short article on "excuses" in *Encyclopedia of Ethics* (2001) is referred to for a general description, the character and the excuses; it can be asked what kind of reception these get in the theories of responsibility that were discussed here. Four examples regarding their acceptance or rejection are looked at as a cross-reference. One or the other is usually foregrounded. According to her, *excuse* as she understands it refers to the total variety of modification of the reactive sentiments. As a description of the moderate modification of the reactive attitudes she uses the term *mitigating circumstances*. In the end, it is advocated that Watson's and Wallace's way of using the concepts should be used here.

pragmatic use. What is important about Doris is that he usefully criticizes character-centered approaches to ethics, and points out the fallacy of using the psychological models of ancient Greece.

- Kelly's critique of the excuses

Erin I. Kelly has written an article in the recent book on "Blame" by Coates and Tognazzini (2013), in which she discusses the imprecise nature of excluding agents from moral responsibility assessments based on excuses related to the level of moral capability of those agents. Her own amendments to the use of the concept of excuse are looked at after the general presentation of the concept.

- Examples

Finally, the ways that the concepts are utilized in the theories of moral responsibility discussed above are demonstrated using examples of situations in which an extraordinary agent (or an extraordinary condition of one) affects the outcomes responsibility assessments. The examples discussed in chapter 6.3 include cases where the agent is a child and one where the agent belongs to the group referred to as psychopaths. Looking at these examples demonstrates the points at which uses of the two concepts intersect.

6.1 MORAL CHARACTER

Moral character, used as a description of the relatively permanent psychological features of the human mind, has a long history. The word *character* originates from philosophy of the Socratic philosophers as well as the Stoics. The author of the *Stanford Encyclopedia of Philosophy* article on “Moral Character”, Marcia Homiak (2015) describes the manifold events in the history of the concept. Starting as a derived word from ancient Greek *charaktêr*, which meant the impression made on a coin, has evolved toward the description of a distinct quality or individuality, and has come to be used almost synonymously with “personality”. (Homiak 2015.)

Although Aristotle’s theory of moral character is the most frequently cited, the concept does not necessarily relate specifically to his teachings. Different varieties of character theory have appeared in at least the works of David Hume and Adam Smith, Immanuel Kant, John Stuart Mill and Karl Marx. In the current philosophical discussion, moral character has made a comeback following Anscombe’s article “Modern Moral Philosophy” (1958). The notion that philosophers should discuss a concept such as character has also come under attack from social psychology. Additionally, epistemological concerns have affected the way character is being discussed more recently. According to Homiak there are at least two basic meanings for character. The first is based on how the term was understood by the ancient Greeks (Homiak 2015). Generally speaking, this philosophical view of character discusses virtue on many levels; on the one hand whether it is based on the cognitive, affective aspects of mind or both, on the other in pragmatic and political terms how virtue should be educated and what virtues are required from rulers. More recently, an alternative way to understand character has been popularized, mainly by modern psychology. This modern conception of character is roughly equivalent to the concept of personality. (Ibid.)

It is argued here that the current criticism of character has more to do with the latter understanding than the former. What follows is a summary on Aristotle’s views on moral character that are relevant in the context of moral responsibility. The Aristotelian notion of character, introduced in chapter 5.1.2, is described in Homiak’s article as being embedded in the linguistic history of the use of the word “character”. Aristotle’s moral excellences were more accurately described in the original texts as excellences of the character (*êthikai aretai*). Therefore, when we talk about moral virtue in his sense, we are not referring merely to individuality but the assemblage of admirable traits that constitute the sort of person we are evaluating. (Ibid.)

According to Aristotle’s theory, the actions of a moral agent are guided by character traits such as virtues and vices.²³⁴ They are a combination of

²³⁴ “Excellence” is a common term used synonymously with virtue as was discussed by H. Smith (2001). See also ch. 5.1.2.

natural abilities and dispositions that define the human personality. Thus, character is an umbrella term for a wide variety of psychological features that have been notoriously difficult for philosophers to accommodate, especially in the theories inclined toward normativity.

Book III of Aristotle's *Nicomachean Ethics* contains his writings on the subject of moral responsibility. As the subject was explained in the previous chapter (5), his writings can be discussed in terms of responsibility for action, as well as responsibility for character. In addition, Aristotle's political responsibility can be interpreted as belonging to the same whole, but whether the notions dealt with in *Politics* can be considered as part of moral responsibility can be problematic.²³⁵ It was stated that when considered together, responsibility for action and responsibility for character could be considered the components of Aristotle's "big picture" of moral responsibility (see Meyer 1993, 122).

Just as Aristotle's ethical theory is interwoven with his political theory, the particular concept of character involves extensions to the political realm. Homiak writes that Aristotle sees achieving the rational capacity in full as requiring broader social relations. By definition, people are naturally political and we need a political community to realize that capacity. (Homiak 2015.) The core statement on responsibility for action occurs at the beginning of book III in *Nicomachean Ethics* (NE III 1, 1109b30–1109b33), but perhaps a more descriptive account is found in the supposedly earlier work, the *Eudemian Ethics* (EE 1223a9–16). Aristotle's ideas on responsibility for character are in book III.5. Although most scholars focus on his idea of voluntariness, which is a key condition for his conception of responsibility for action²³⁶, book III of the *Nicomachean Ethics* also contains his thoughts on the alternative: "[I]t is irrational to suppose that a man who acts unjustly does not wish to be unjust or a man who acts self-indulgently to be self-indulgent." (NE III.5., 1114a 8–14.) The unjust or self-indulgent man acts as his character dictates throughout the act.

Virtues do not appear out of nowhere, but form an integral part of how societies function and how habits and customs pass from one generation to another. Homiak discusses how the education of character happens in the *Nicomachean Ethics*, taking both the emotional responses and the rational

²³⁵ In the book's context, responsibility can extend through character into the political sphere, but as no further idea other than the agents "membership of human communities" exists in the Strawsonian context, comparing these two strands of thought would be difficult (Strawson 2008, 17). For this reason, comparison between the political responsibilities of the two theories is not pursued here further. While the previous chapter ruled out the topic of political responsibility, we saw that this aspect would be referred to when relevant. Otherwise, primarily in Aristotle's case the two sides of moral responsibility (for action and for character) are regarded as the main topics of interest.

²³⁶ T. H. Irwin (1980) is used here on another occasion to sketch out Aristotle's responsibility for action. His central summation of Aristotle's theory of action follows: "A is responsible (a proper candidate for praise and blame) for doing x if and only if A does x voluntarily" (Irwin 1980, 125.)

assessments into account, in that the former require the right kind of character development. Education leads to correct attunement toward what we take pleasure or pain from. *Phronēsis*, or practical wisdom, is the faculty that directs action to avoid acts associated with vices and toward pursuit of those considered virtuous. (Homiak 2015.)

The classic description of virtues and their situation between two corresponding vices can be found in *Eudemian Ethics* book II.3. In Aristotle's moral theory, the golden mean, which signifies a mean between the two extremes of vices related to a virtue, is the ideal for the virtues in most cases. The central point between the complete lack and excess of the virtue is the desired place. For example, for courage, the lack of the virtue would be called cowardice and where bravery goes too far, foolhardiness would be the vice. (*EE* II.3. 1221a.)

As discussed by Richard Sorabji (1980), Myles F. Burnyeat (1999) and Nancy Sherman (1999), Aristotle's virtues are constituted by examples from habituation and education. From early childhood to young adulthood, the agent is receptive to outside influences in adopting patterns of behavior that create the stable, fully-formed virtues. Practical reason (*phronēsis*) has an important role in how well the character virtues are adopted, but it does not determine their content (see Sorabji 1980, 210). Character virtues are formed instead by practice. An agent's character can certainly be affected by training; however, it is not in itself enough to create a virtuous character (Sorabji 1980, 211). Aristotle follows Plato's notion that the individual could form habitual virtues through upbringing (Sorabji 1980, 211–212). By initially practicing the habit virtue, the character virtues become what the action of the individual determines and thus become virtues proper. The origin of character virtues resides in the society as well (see Broadie 1991, 170).

Education and the city state are essential in generating the intellectual virtues. Practical reason (*phronēsis*) is an intellectual virtue. Practical reason incorporates the act of deliberation and reflection, leading ultimately to the selection of the act that fulfils the end, called *prohairesis* (Broadie 1991, 179).²³⁷ The concept is so crucial in *Nicomachean Ethics* that it is considered the defining trait of human action (NE VI.2., 1139b5–6). The deliberated choice reveals the virtues or vices of the actor to the society. In this case, praise or blame is given on account of these states of character.

Virtues affect actions because in determining responsibility, the virtues (or excellences)²³⁸ appear as a guiding property of character. They decisively affect selection of the act. The act is selected from all possible alternative

²³⁷ The word was discussed in chapter 4, meaning “a choice made with full awareness of alternate possibilities” (Salkever 1990, 69).

²³⁸ According to Aristotle, the soul consists of three different things: 1. passions (*pathos*) 2. faculties (*dynamis*) and 3. states of character (*hexeis*). The states of character are essentially virtues (*arête*) and their accompanying vices (*kakia*) and his ethical theory of the golden mean discusses their relations. (See NE II, 5.)

choices leading to the same end. The behavior of the actor is determined by causes originating from herself. Nevertheless, no one can become virtuous in a vacuum, since the virtues originate in society. To demonstrate this, Aristotle makes a distinction between intellectual virtues and character virtues. (*NE* II.1, 1103a15–21.)

Like examples used on other occasions, Aristotle alludes to the cases of animals, children and the mentally ill. His theory postulates that only adult human beings are capable of deliberated, voluntary acts, which (2b) excludes actors belonging to the groups mentioned (see *NE* III.2, 1111b4–9).

The effect of character on the act in Aristotle can be also described as follows: the concept of character in *Nicomachean Ethics* as it relates to moral responsibility signifies the habituated conduct provided by the agent's upbringing and education within her society. Its constitutive elements, the virtues, describe the state of that person's character and illustrate how that person is likely to act in different situations. They can be interpreted as a part of the decisive mental faculties that either select the act from among the possible courses of action, or if there is only one course, enforce its commission or omit it. As was noted before, such claims for voluntariness and historicity are part and parcel of the incompatibilist "double argument" for free will. Watson's example questioned this relation, and Fischer & Ravizza provide a theory that lets these two ideas co-exist: as Eshleman noted, the theory of these authors falls into the categories of "merit-based" and "compatibilism", which is breaking the norm that merit-based responsibility is typical of the incompatibilists (Eshleman 2014/[2001]).

Discussion of Aristotle's excuses normally mentions the main requirements for moral responsibility in his theory of action. These were discussed in connection with Irwin's simple theory (5.1.1), and include the two excuses related to voluntariness: ignorance and coercion. (Broadie 1991, 142-148.) Originating as a criticism of Plato's claim that no one knowingly does wrong, Aristotle grouped acts for which the actor or the agent could be responsible into two categories (*NE* III.1.). First, the act can be done in full knowledge and, second, it can be done voluntarily. These two conditions allow for multiple situations in which the act is deemed to be wrong.

Voluntariness and character affect each other. At the moment the act takes place, it's possible that the unjust or self-indulgent act is done voluntarily and in full knowledge. One can do wrong, and can be held responsible for the act and be blamed for it, even if the actor is acting voluntarily, and in full knowledge of the situation. (Roberts 1989, 32-33.) According to Aristotle, the actor is simultaneously responsible for his actions directly and indirectly through his character.

In Aristotle's theory of responsibility, the excuses referred to (ignorance and coercion) are not necessarily related to character. Instead, in terms of responsibility both the considerations related to these excuses and the questions related to character are relevant in considering *all* related factors ,

whether the agent is blameworthy or not. Which factors are required to determine a sufficient level of moral responsibility is being debated.

Skepticism of character

Philosophical theories relying on moral character have recently had a rude awakening in the wake of John Doris's and Gilbert Harman's work, which has set out to heavily criticize the use of character as having no basis in human psychology and social psychology.²³⁹ Homiak criticizes the attack by social psychologists as falsely focusing on "local" traits that do not include the complete behavioral patterns that moral philosophy is otherwise capable of discerning. (Homiak 2015.)

Another commentator, Mark Alfano (2013a), has summarized the current debate as well as listing its participants²⁴⁰ calling the topic "Scepticism of Character". Alfano clarifies the situation of the debate by distinguishing three reactions to the attack:

*Skepticism about character draws on research in the social sciences (especially social psychology in the situationist paradigm) to argue that both folk psychological and neo-Aristotelian conceptions of character as a robust disposition to act on reasons is empirically inadequate. In John Doris's formulation, traditional character-based ethical views tend to be globalist in that they are committed to cross-situational consistency of character, temporal stability of traits, and evaluative integration of virtues within individuals. Psychological research suggests, however, that while traits are stable, they tend to have low cross-situational consistency. This means that someone who is stably disposed not to lie to his friends may nevertheless be quite prone to stealing from his relatives or cheating on his wife. Talk of broadly individuated virtues such as honesty might therefore seem empirically inadequate. Virtue ethicists have replied in a variety of ways. The three primary responses are (**the dodge**) virtue ethics is primarily a normative theory, so we should be untroubled if most people are not virtuous; (**the retreat**) the commitments of virtue*

²³⁹ Thanks to Markus Neuvonen for bringing my attention to the topic of skepticism of character.

²⁴⁰ Alfano writes on philpapers.org on "Scepticism about Character" (2013b): "John Doris's original paper on this topic is Doris 1998, which was quickly superseded by Doris 2002. Almost at the same time, Gilbert Harman developed his own variant of skepticism about character in Harman 1999, 2000, 2003, and 2009. Advocates of the dodge include Badhwar 2009 and Kupperman 2009. Advocates of the retreat include Hurka 2006, Merritt 2000, and Miller 2009. Advocates of the counterattack include Kamtekar 2004, Russell 2009, and Sreenivasan 2002. Other related research, such as Sarkissian 2010 and Alfano 2013, has attempted to move past the question of empirical adequacy by focusing on ways in which situations can be modified or framed to promote moral conduct." (Alfano's bibliography on philpapers.org based on Alfano 2013a.)

ethics can be weakened in such a way that it becomes empirically adequate; and (the counterattack) the evidence from social psychology does not in fact challenge the virtue ethical conception of character. (Alfano's bibliography on philpapers.org based on Alfano 2013.)

As the modern debate on responsibility resumed with Strawson, a return to the classical sentimentalism was anticipated but not actually achieved until the turn of the millennium. While virtue ethics correctly highlighted where the default conception of responsibility was lacking, there have been opinions voiced that virtue ethics failed to describe character properly in modern terms. The antiquated psychology of Aristotle has served as a tool providing an alternative way to conceptualize both responsibility and character, but it has often been at odds with the discoveries of modern neuroscience and other psychological experiments. Doris's *Lack of Character* (2002), relying on Strawson and Wallace, attacks the conception of character as a philosophical, permanently established feature.²⁴¹

John M. Doris' *Lack of Character* (2002) is a fairly recent work of great interest for study of the conceptual relations between responsibility²⁴² and character. He highlights the problem that over-emphasizing character means essentially diluting any opportunity for responsibility assessments taking place immediately after the act in question. In this way, he fits quite comfortably into the Strawsonian mindset in supporting reactive attitudes as having more weight instead of the possible historical events and circumstances that led to the act.²⁴³ Doris writes quoting Wallace, but in a way also relating to Bernard Williams's idea that a theory of moral responsibility should not be characterological, as he thinks that character is ineffective in describing actual psychology and that everyday assessments of responsibility do not have consistent connections with any idea of character. (Doris 2002, 130.)

He concludes that:

Thinking in terms of character, insofar as it accurately "summarizes" previous behavioral trends, may serve an epistemological role in responsibility assessment; it highlights behavioral anomalies that may fall under exempting or excusing conditions. But this summary use of characterological discourse need not invoke the conception of character I have argued against, since it makes no commitment to

²⁴¹ Gilbert Harman (2000) wrote almost at same time as Doris, whose first publication is from 1999. At a more general level, these studies at the meeting point between moral psychology and more empirical disciplines have resulted in interesting discussions on many subjects (see Prinz 2007, Sinnott-Armstrong (ed.) 2008, et al.).

²⁴² Doris has a Strawsonian conception of responsibility. He also refers to Watson's and Wolf's work. (Doris 2002, 128-131.)

²⁴³ An "acharacterological account of responsibility need not be apsychological" Doris 2002, 129

robust and evaluatively consistent dispositional structures (Doris 2002, 131).

This highlights the strength of Doris's position. According to him, character can be useful as an everyday summary of psychologically relevant past events. This instrument of character can help people to quickly evaluate each other's behavior. Noted, in this way Doris does admit that this limited use of character may serve an epistemological role, so that it seems even he, who is most explicitly against the concept of character will not leave it completely out of the theories of responsibility. In his theory, the concept of character is relevant to the assessment of responsibility as a *cognitive tool*.

Recalling Marion Smiley's book (1992), Doris echoes the distinction between causal responsibility and blameworthiness, as he thinks that causal responsibility in the form of "consequentialism" is not enough, but that "psychological state" must also be considered as well in all responsibility assessments. Yet blameworthiness as determined by the psychological connection is insufficient by itself as excuses may be considered (e.g., as in the case of a 5-year old child). (See Doris 2002, 129.)

Doris gravitates toward Wallace's views (1994), as he supports Wallace's ahistoricity. In *Lack of Character* Doris notes the complications by mentioning examples concerning children and "the deranged", but doesn't tackle the subject, as he writes: "we shall be on the right track if we say that responsibility assessment consists in establishing the presence of causal and psychological connections and the absence of excusing and exempting conditions." (Doris 2002, 130.)

The problem with Doris's adoption of the Strawsonian schema is that the "characterological" accounts, which Doris is arguing against, *are* trying to describe and take these "excusing and exempting conditions" into account. Therefore, if Doris requires the absence of the latter in order for responsibility assessments to be possible, he limits the theory as much as Wallace did with the "blameworthiness inhibitors". "Yoking responsibility to character" and falsity of the condition "absence of excusing and exempting conditions" are essentially two sides of the same coin, and describe the paradox of responsibility quite well: you can always dilute responsibility for action too far if you consider every historical aspect leading to the act. At one end of pure causality is excessive abstraction, at the other end of extreme historical effect is hard determinism.

Another problem with Doris is that his Aristotle scholarship is selective at best. While he does refer specifically to *Nicomachean Ethics* book III.1 (NE 1109b30-5; 1111a22-5), he doesn't mention book III.5. at all, which is curious, because that's where Aristotle's characterological "amendment" to his "theory of responsibility" is. Doris advocates, however, T. H. Irwin's attempt to formulate an Aristotelian account centered on deliberative capabilities, an approach he sees as related to his own. (Doris 2002, 212, n7.)

Basically then, Doris is emphasizing the responsibility for the action side of Aristotle's theory while not going into responsibility for character or at least he doesn't do so in depth. Taking Aristotle's character responsibility in *NE* book III.5 into account would be contradictory to Doris's claim that "the progenitor of character ethics is himself not obviously committed to a characterological account of responsibility". While this is true as Aristotle is not "committed" to any account of responsibility, the character responsibility bit in III.5 can be included in his account of the subject if so interpreted.

In short, Doris might be unaware of Aristotle's complete account of responsibility. To be fair, he admits that "I don't much consider the extraordinary scholarship Aristotle inspires" (Doris 2002, 28). This means that while his main targets are the characterological accounts of responsibility (virtue ethics included), which began with Aristotle, he hasn't studied Aristotle himself as much. The ideal case then would be to reconcile Doris's most important contributions, augmented with the more recent moral psychological advances with a deeper understanding of the intricacies of Aristotle's (and Hume's for example) theories. Thus, it can be argued that Doris's account in terms of *philosophical* character is incomplete. For example, he does note with interest Bernard Williams's elements of character, but finds intention (the 2. element) problematic and thus omits it (Doris 2002, 211, n5).

Homiak points out the weakness of the situationist challenge, which she believes lies in misunderstanding the component concept of "trait" between the two disciplines:

[T]he character traits criticized by situationists have little to do with the conception of character associated with the ancient and modern moralists. The objectors say that the situationists rely on an understanding of character traits as isolated and often non-reflective dispositions to behave in stereotypical ways. They wrongly assume that traits can be determined from a single type of behavior stereotypically associated with that trait. [...] In short, the objectors say that the situationists rely on a simplified view of character. They assume that behavior is often sufficient to indicate the presence of a trait of character, and they ignore the other psychological aspects of character (both cognitive and affective) that, for most of the philosophers discussed in this entry, form a more or less consistent and integrated set of beliefs and desires. (Homiak 2015.)

A brief reflection on how character in philosophy is different from character in psychology — essentially the former is more complex than the latter. The former also incorporates the latter but does not limit itself to it. The philosophical notion of character contains the psychological traits of the person, but also includes their past decisions and acts, at least those that are directed toward the agent him/herself.

What is the situation of character as understood by current psychology? In psychology, there is a long drawn out battle between situationism and personality theories that reflects many features of the determinism vs. free will debate. Could the statements produced by the neuropsychologists / experimental philosophers contribute toward the responsibility debate through an updated vision of human intellectual processes?

What is meant by moral character here is the Aristotelian view of character, consisting of virtues and vices. In relation to moral responsibility, these character traits have a persistent effect on the behavior of the agent and thus are important to moral responsibility in the form of blameworthiness. The result is that considerations related to the character of the agent may mitigate the response to the wrong-doing committed by the agent. For example, a child who has an underdeveloped character can be deemed not blameworthy for that reason. On the other hand, the ability of the agent to influence his or her own behavior through manipulating his or her character through training opens the door for moral responsibility to include notions of historical responsibility within responsibility assessments.

The recent attack on character starting with the work of John Doris (2002) appears to partially miss its mark in having to do mostly with the contemporary concept of character, which is synonymous with personality. As such the criticism, while very welcome in reinvigorating talk on character, is aimed more at the psychological conception. Therefore recalling Alfano's strategies(2013), criticism of character is here largely "dodged", but it is noted that the general momentum of the current debate should not be wasted, and that even more viable theories, relying on current empirical knowledge of the underpinnings of character should be pursued. Just as Doris concedes that character is needed as a cognitive tool in everyday situations, it is necessary to go one step further and keep the moral excellences as basic vocabulary for the reasons that Watson referred to as *aretaic appraisal* (see chapter 4.3.4).²⁴⁴

This concludes the summary of what is referred to as character here, and it will now be followed by an enquiry into what the definitions excuse are, and which of these is followed here.

²⁴⁴ For example, if it were conceded that the results of Milgram's tests would compel us to exchange our notions of character for specific descriptors of psychological functions in education. In this case, the imagined encouragement toward justice, courage or any other appraisable goal would become too complex for kids especially to understand. If teachers encouraged pupils with certain temperaments to hold back or try harder, the reason and explanation of the encouragement in situationist terms would require infinite justifications.

6.2 MORAL EXCUSES: FROM STRAWSON'S SPECIAL CONSIDERATIONS TO WATSON'S EXEMPTIONS

This chapter describes four different renditions of the same theme. First, J. L. Austin's "A Plea for Excuses", the article in which the study of excuses is described as its own "ramiculated branch" of philosophy.²⁴⁵ Second, the examination moves to *FR*, which has already been discussed, but is considered so essential as to demand further scrutiny. According to Strawson, the reactive attitudes are subject to possible alterations in the form of certain special considerations or pleas. These considerations can modify the initial reaction experienced by a morally qualified person. How precisely this takes place will be discussed below.

The Strawsonian use of special conditions and pleas has not been employed as a practice by his followers, but the extensions of these words explored by various writers. Gary Watson has distinguished between excuses and exemptions in his article "Responsibility and the Limits of Evil" (Watson 1987), a practice followed here as well. These concepts are also used by R. Jay Wallace (in Wallace 1994), citing Watson's article on this issue. It is stated at the end of this chapter that Watson's and Wallace's version is the preferred one.

Finally, Holly M. Smith's distinction between *mitigating circumstances* and *excuses* (H. Smith 2001, 506-507) is discussed as an alternative way of labeling the two main varieties of special conditions. Smith's mitigating circumstances and excuses appear to be directly comparable with Watson's and Wallace's use of the excuses and exemptions. The meaning of "excuse" is the notable difference between the interpretations in that where Watson and Wallace use the word to mean the partial suspension of the reactive attitudes, Smith means the their total elimination, which is referred to as "exemptions" by the previous authors.

The origin of the use of excuse as a philosophical instrument in the current debate can be traced back at least to J. L. Austin, who writes in his "A Plea for Excuses" (1956/57) about the difference between justifications and excuses, claiming that what is distinctive about the excuse is that it concerns an act that has led to wrong-doing or harm. Excusing a well-meaning,

²⁴⁵ Austin's major article on excuses, "A Plea for Excuses" (1956/7), can be read as a forerunner to the particular subject of the philosophy of responsibility. By sharing the academic circle and being the senior scholar to Strawson, his influence is on the latter author is undeniable. Being situated deeper in the more language-based discipline of the ordinary language philosophy, there is some trouble in comparing the ideas with the way of doing philosophy that prevails today. The same is true in Strawson's case as well. Nevertheless, the discussions about the relevant concepts and similarities with the struggles of the earlier attempts to tackle the same subjects are always of interest. Hence Austin is mentioned here as a reference point.

benevolent act with universally acknowledged good consequences would be pointless. *Excuses* according to Austin are not a concept, or even a theme in philosophy, but “the name of a whole branch, even a ramificated branch, of philosophy, or at least of one fashion of philosophy.” (Austin 1956/57, 1.) This implies that the study of the subject had become a small sub-discipline of philosophical study. The short description of Austin’s excuse reads as follows:

In general, the situation is one where someone is accused of having done something, or (if that will keep it any clearer) where someone is said to have done something which is bad, wrong, inept, unwelcome, or in some other of the numerous possible ways untoward. Thereupon he, or someone on his behalf, will try to defend his conduct or to get him out of it. (Austin 1956-57, 2.)

Excuses thus, as conceptualized by Austin, are closely related to responsibility: ²⁴⁶ he writes that for A to commit any X presents the agent with the possibility to accept responsibility for the act. In addition to merely accepting responsibility, the agent can also agree that the act was bad or disagree with it. In this way, Austin keeps accepting responsibility and viewing one’s action as bad separate. In other words, it is possible for the agent to accept responsibility for the act, but simultaneously deny that the act was bad in itself. (Austin 1956/57, 3.)

Excuses as a group of statements are initially distinguished from *justifications* in Austin’s article. The concepts are close enough that they are often confused with each other. While these might have some overlap, one trait of the excuses seems clear: an excuse seems to imply wrong-doing in a way that a justification does not necessarily do. There can thus be said to be a connection between an excuse and the mitigation of blame (Ibid. 3).

As Austin describes, doing an action in philosophy is “a highly abstract expression” and for this reason excuse can function as a stand-in for practically any verb. This does not mean, however, that “not every excuse is apt with every verb”. The first problem in classifying excuses, Austin argues, is related to the question “What action is free?” He continues by giving the example of Aristotle, who has been criticized for his apparent lack of recognition of “the real problem” of determinism and free will on many occasions. Related to this, Austin states that these attitudes toward Aristotle’s work and the charge of neglecting the problem of determinism have led Austin to the question of the nature of the excuses. (Ibid. 6).

As for the connection between the concepts of moral responsibility and the excuses, Austin thinks that *responsibility* would serve better than

²⁴⁶ As a specific list of which conditions qualify for excuses, Austin mentions, examples of: mistake, culpable negligence, accident, inadvertence and “automatical” actions as such that call for an excuse (see Austin 1956/57, 21-24).

freedom in discussing the problem of determinism; thus Austin essentially frames the development that Strawson's paper put really in motion. Nevertheless, avoiding enthusiasm about the word, Austin does note the limitations of the concept of responsibility: "But in fact "responsibility" too seems not really apt in all cases. I do not exactly evade responsibility when I plead clumsiness or tactlessness, nor, often, when I plead that I only did it in the circumstance of no choice". (Ibid. 7.)

Austin states that, being under a constraint does not rule out accepting responsibility. This implies a difference between accepting responsibility and being held responsible, which is not apparent in Austin's case. We may summarize, by saying that that Austin believes that the clarification of responsibility can be helped by investigating the excuses (ibid). He provides a number of reasons why the rubric of excuses is worth exploring. Without going deeply into these here,²⁴⁷ Austin emphasizes the need for philosophers to treat words as clear, carefully-defined tools. After all, he did follow the ideals of early 20th century analytical philosophy. (See ibid. 7-9.)

As pointed out in chapter 3.1.3, Strawson's type-1 and type-2 pleas were of especial interest to this work.²⁴⁸ Strawson does not actually use the concept of *excuse*, but his discussion about the special considerations and pleas fits the bill (*FR*, 7-8).²⁴⁹ While Strawson does not quote Austin directly, he did work in the same academic environment as Austin.²⁵⁰ His text is clearly influenced by Austin's work. For example, the pleas that Strawson lists seem similar to those in Austin's article.

Where character was discussed in the Aristotelian context, the Strawsonian context seems to deal with the same subject (of some property

²⁴⁷ This is mainly because Austin is doing a particular kind of linguistic philosophy, which it is not possible to explain here in style or quality.

²⁴⁸ Strawson himself does not use the "type-1" and "type-2" distinctions. These are used by Gary Watson in his "Responsibility and the Limits of Evil" (1987, 260) in order to clarify Strawson's concepts.

²⁴⁹ As was mentioned above, Strawson does not systematically use the word "excuse" in *FR*. However, the sections where he does speak about special considerations or pleas are clearly set apart and the followers of Strawson (namely Watson, Wallace, Wolf and Darwall) consistently refer to these topics using "excuse". Thus, it is fairly clear what an excuse generally is in these authors. (See *FR*, 7-11.) As for the specific meaning, there are a couple of interpretations (Watson's and Wallace's as well as Holly Smith's), which are looked at later, and one of these interpretations is advocated here.

²⁵⁰ Strawson also followed Austin in that both belong to the so-called ordinary language school of Oxford philosophers. This way of doing philosophy followed Wittgenstein's second phase and the work of the Vienna Circle and lasted in Oxford roughly from 1945 to 1970. The main philosophical leanings of this group were linguistic and the way its members pursued philosophy was to prefer the "use" of the words over their "meaning". Critics of ordinary language claimed that the discipline deliberately conflated semantics with pragmatics, and ultimately the project shut down. As one of the ordinary language philosophers, Strawson can be read as having influenced contemporary philosophy in being more focused on *concepts* than *language*. (See Parker-Ryan 2012.)

related to the agent's status affecting the response to her wrong-doing) in terms discussed by Strawson and referred here to as excuses and exemptions after Watson (1987) and Wallace (1994). The former coined the distinction between excuses and exemptions (Watson 1987, 259-261), and Wallace adopted his terms (Wallace 1994, 118). Wallace also refers to Austin's distinction between justifications and excuses, an excuse implying that the act that the response is related to was wrong regardless of whether the agent is blameworthy or not (see Austin (1956-57, 23-24); Wallace 1994, 119-120).

In the most common case, blame boils down to the sentiment of *resentment*, directed at a *wrong-doer*. In Coates & Tognazzini's introductory article to their anthology *Blame* (2013), the Strawsonian approach to blame is clearly defined: "to blame is to target [the object] with one of the reactive sentiments" (Coates & Tognazzini 2013, 13). The Strawsonian account expands from this simple premise, and all is well when the agent and the situation follow the norm. But as one wants to utilize reactive sentiments in a broader variety of considerations, a problem arises: What about the exceptions to the norm? According to Strawson, the cases that deviate from the usual or the "normal" agents are situations in which the original and spontaneous, reactive sentiments may be subject to alteration.²⁵¹

Strawson lists two groups, one with two sub-groups, so that altogether there are three distinct categories of situation in which excuses are applied in order to curtail the natural reactive sentiment that, without the property in question, would otherwise lead to blame. In actuality, the agent is rarely identical between each case. Here we consider these special situations in which reactive sentiments are modified or "mollified" or removed altogether, because of certain properties related to the agent. (*FR*, 7.)

Strawson writes on the special considerations:

To the first group belong all those which might give occasion for the employment of such expressions as 'He didn't mean to', 'He hadn't realized', 'He didn't know'; and also all those which might give occasion for the use of the phrase 'He couldn't help it', when this is supported by such phrases as 'He was pushed', 'He had to do it', 'It was the only way', 'They left him no alternative', etc. [...] they have something still more important in common. None of them invites us to suspend towards the agent, either at the time of his action or in general, our ordinary reactive attitudes. They do not invite us to view the agent as one in respect of whom these attitudes are in any way inappropriate. They invite us to view the injury as one in respect of which a particular one of these attitudes is inappropriate. They do not invite us to see the agent as other than a fully responsible agent.

²⁵¹ "It is one thing to ask about the general causes of these reactive attitudes [...] it is another to ask about the variations to which they are subject, the particular conditions in which they do or do not seem natural or reasonable or appropriate" (Strawson 2008, 9).

They invite us to see the injury as one for which he was not fully, or at all, responsible. (FR, 7-8. My emphasis.)

The first group of excusing conditions is classified by being somehow unintentional or such that an outside influence forces the agent to take a certain course of action. It could be labeled “unintended consequences and coerced acts” for example. Strawson lists a number of expressions that have been employed in these situations: “He didn’t mean to”, “He hadn’t realized” and “He didn’t know”, and so on exemplify the cases in which the agent causes something unintended, while “He was pushed”, “He had to do it” “it was the only way” or “They left him no alternative” are situations which involve some external coercion. Common to these examples is the possibility that the agent is a fully developed, functional human being. The agent, while causally responsible, may be subject to mitigated blame, as being in full knowledge and capability; she would have done otherwise, but if this is done it is not because of the quality or state of the agent herself. (Ibid. 7-8.)

Strawson continues on the second group of special considerations:

The second group of considerations is very different. [...] we may think of such statements as ‘He wasn’t himself’, ‘He has been under very great strain recently’, ‘He was acting under post-hypnotic suggestion’; in connection with the second, we may think of ‘He’s only a child’, ‘He’s a hopeless schizophrenic’, ‘His mind has been systematically perverted’, ‘That’s purely compulsive behaviour on his part’. Such pleas as these do, as pleas of my first general group do not, invite us to suspend our ordinary reactive attitudes towards the agent, either at the time of his action or all the time. [...] They invite us to view the agent himself in a different light from the light in which we should normally view one who has acted as he has acted. I shall not linger over the first subgroup of cases. [...] The second and more important subgroup of cases allows that the circumstances were normal, but presents the agent as psychologically abnormal — or as morally undeveloped. The agent was himself; but he is warped or deranged, neurotic or just a child. When we see someone in such a light as this, all our reactive attitudes tend to be profoundly modified. (FR, 8-9. My emphasis.)

The second group contains two sub-groups: as stated in the third chapter, the first subgroup, which was of less philosophical interest than the second, and the second subgroup that in turn sees the agent himself/herself as “abnormal” or psychologically disturbed. What is very puzzling is that this group also includes children. Common to these statements in the second group is that they describe situations that call the normal rules by which we evaluate each other into question. The reactive sentiments that would ordinarily be directed toward the agent are reconsidered on basis of some additional information about the agent in question. (Ibid. 8.)

Thus, as for the excusing conditions, Strawson lists three possible alternatives: (1) *unintended consequence or coerced acts*, which are exemplified by “He didn’t mean to”, etc. The second group is divided into two subgroups, the latter more interesting: (2a) *altered state through circumstances* being described as situations in which, the agent is, for internal or external reasons, acting in an out-of-the-ordinary fashion and, finally, (2b) *some altered state through psychological abnormalities*. It is noteworthy that Strawson includes children within this last group, but he remains careful to refer to them as their own separate group, with insufficient development being the greatest cause for exempting them.

As discussed in 3.1.3, the fourth part of *FR* introduces three types of special excusing considerations that can modify, mollify, or even cancel out the reactive sentiments (*FR*, 7-10). The reactive sentiments, which were thought by Strawson to be the constitutive manifestations of the responsibility relations in moral communities were the direct response to an act in the examples that led to injury or comparable harm.

Reactive sentiments or emotions, feelings or attitudes refer to a set of emotions that arise spontaneously when an act is observed, and through these emotions a sort of basis for moral judgments is formed.²⁵² Strawson writes of what qualify as reactive sentiments: “resentment and gratitude [as a] usefully opposed pair. But, of course, there is a whole continuum of reactive attitudes and feeling stretching on both sides of these and – the most comfortable area – in between them”. (*FR*, 7.)

The excuses affect the reactive emotions by ameliorating the initial sentiments or even completely negating them in certain circumstances. There are two ways that this might happen: the first is that the person’s behavior may be excused, for instance, if it is determined that the act committed was as an accident. In the other case, the participant attitude may be given up entirely by being switched to the objective standpoint. The individual in question is then no longer viewed as a fully competent member of the moral community.²⁵³

The other popular subject for examples is those individuals who are impaired psychologically for any reason.²⁵⁴ The psychopath is a unique case which fits into this group, wherein the person might fully understand what moral requirements the community expects, but sees no reason to follow them and does not care to. Even worse, the psychopath might use these expectations in cruel feats of manipulation.

Strawson describes the example in which someone treads on his hand, arguing that how we react to the act carries a lot of meaning: if the act was

²⁵² Strawson harks back to the moral sense theories and their followers, the classical sentimentalists, the most renowned being David Hume and Adam Smith (see Strawson 2008, 25-26).

²⁵³ The word *exemption* is used in connection with these forms of “full pardon” below following Watson (1987) and Wallace (1994).

²⁵⁴ See P. F. Strawson 1962: 6–10; Watson 1987: 259–260; R. Jay Wallace: chs. 5–6.

done on purpose we are likely to get angrier than if it was an accident. Strawson briefly notes that forgiveness²⁵⁵ has received less attention than the more negative reactive emotions of resentment, indignation and guilt. (See Strawson 2008, 6.)

While Austin drew up a picture of excuses as a complete field of study in philosophy, there are writers who attempt to define the word as a singular concept. While Strawson did not actually use the concept of *excuse*, this word did apply to “special considerations” or “pleas” (*FR*, 7–8). As stated before,²⁵⁶ Gary Watson made the comparison between the different groups of Strawson’s special considerations and labeled the type-1 pleas *excuses*, as distinct from cases with type-2 pleas or the *exemptions*. This distinction was so appealing that Wallace adopted it in his *Responsibility and the Moral Sentiments* (1994). Wallace’s description of these groups is discussed in this section for clarity’s sake.

Wallace, whose account of responsibility is here considered to be the contemporary variant of the Strawsonian theory of responsibility (with Wallace’s own, Kant-inclined additions). As was noted in chapter 4.1, Wallace is discussed here as his use of blame is more articulated than is offered in Strawson’s essay and because the connection between the concepts of blame and responsibility is clearer in it. (See Coates & Tognazzini 2013, 13–15.)

As implied above, both Strawson and Wallace are part of J. L. Austin’s legacy. Wallace cites Austin: “[T]here is a connection between doing x intentionally and being morally responsible... [So] doing x intentionally is ordinarily a condition for responsibility for x, such that if s has not done x intentionally, then it is unfair to hold s morally responsible for x.” (Wallace 1994, 122.) Responding to this, Wallace asks: “Why would it be unfair to hold s responsible for x if s has not intentionally done x in the first place?” (Ibid.) Arguing against the incompatibilists, Wallace calls for an account of excuses that includes a description of how normativity functions in their case. Thus intention is an important part of action considered morally responsible in Wallace’s context.

Conversely, Wallace is not interested in praise at all, focusing on blame and other negative emotions, as he thinks that they are more constitutive in how the responsibility-relations form. He writes that “praise does not seem to have the central, defining role that blame and moral sanction occupy in our practice of assigning moral responsibility” (Wallace 1994, 61).²⁵⁷ In the following footnote, he extends this to apply to the existing social practice as well, although it is not clear what this practice is. He thinks that the most important emotions are expressed in connection with moral situations, and

²⁵⁵ For example, Garrath Williams has returned to the subject of forgiveness in responsibility in his latest article *Sharing Responsibility and Holding Responsible* (G. Williams 2013).

²⁵⁶ See chs. 4.1.2 and 4.3.4.

²⁵⁷ Compare this with Coates & Tognazzini 2013, 4–5.

that they require responsibility assessment. The emotions included are resentment, indignation and guilt (ibid. 67).²⁵⁸ The most significant, if less obvious cut that Wallace makes is that he almost exclusively talks about *holding* someone responsible.²⁵⁹ He separates other aspects of responsibility into the Kantian idea of *conditions of autonomy* and expands Strawson's two categories of excusing conditions into a slightly different ordering of excuses and a new category of *exemptions*. (Ibid. chapters 5 & 6 respectively.) The excuses are divided into four categories (Wallace 1994, 120-147), while the exemptions list six problematic types of states for moral agents to be in (ibid. 166-180).²⁶⁰

Wallace notes that Strawson's account of excuses addresses the issue that while intention is important for holding someone responsible it is not sufficient to simply check whether one intended an act and be done with it. He continues:

excusing conditions [...] seem to operate locally: they make it inappropriate to hold an agent responsible for a particular action, but they do not make it inappropriate to view the agent as morally responsible in general. Strawson then proceeds to account for the force of the excusing conditions, so understood, in terms of the reactive attitudes. Given the connection between responsibility and the reactive attitudes, a condition will make it inappropriate to hold an agent responsible for a given action if the condition will make it inappropriate to respond to the action with one of the reactive emotions. (Ibid. 124-125.)

Wallace's own expanded account, or his typology of excuses in the chapter "Blameworthiness and Excuses", includes the following four types of excuse: a) inadvertence, mistake or accident, b) unintentional bodily movement, c) physical constraint and d) coercion, necessity and duress (ibid. 136-147).

How do inadvertence, mistakes and accidents differ from one another? At least in Wallace's case, inadvertence means deciding to do x, a course of action that leads to doing something unwanted in the process. In the example, Wallace is getting something from the fridge and then steps on s's

²⁵⁸ Also, in his recent article, "Rightness and Responsibility", (which discusses rightness from the internalist position; another topic not discussed here), included in Coates & Tognazzini 2013, he still uses the trio of resentment, indignation and guilt as the relevant moral sentiments.

²⁵⁹ Wallace 1994, 16. For an example of why exactly this is problematic see Angela Smith's *On Being and Holding Responsible* (2004). Note also that: "Another dichotomy that that has attracted critical comment concerns the gap between holding and being responsible. Critics such as John Martin Fischer and Mark Ravizza, have argued that Strawson [too] is insufficiently sensitive to this gap and that this vitiates much of his strategy in "Freedom and Resentment"." (McKenna & Russell 2007, 12).

²⁶⁰ The list of these categories was recounted in chapter 4.1.2. Excuses were related to accidents, involuntary bodily movements, physical constraint and coercion, and the exempted groups included children, the mentally ill, addicts, psychopaths, the hypnotized and the stressed.

hand. A mistake, in turn, means that *x* is something that is done to the other on purpose, but is mistakenly aimed at the wrong object. A nearby thief could need his hand stepped on, but *s*'s hand gets trampled by mistake. An accident means that the agent might be aware of the wrong-doing, but for other reasons he ends up doing it anyway. Wallace's example of the last case is stepping on *s*'s hand because he is trying to put out the flames of a fire, not intending to cause harm. (Ibid. 136-137).

Another type of action that might be considered to be excused is those where the agent causes something, not actually by doing anything, but by the unintended movements of one's own body. These movements could be caused by acute illness, or some outside force. Wallace mentions tickling, and other more serious situations (ibid. 140-141). Both a) and b) are not sufficient excuses, according to Wallace, if they are the product of gross negligence on the part of the agent herself. Physical constraint can be either external, such as someone grabbing the agent or knocking him out, or internal as in being paralyzed. So, in other words the case of physical constraint means simply that the agent is unable to address what is required, because he or she is prevented or stopped for one reason or another (ibid. 141). Coercion in Wallace's case differs from Frankfurt's theory, as the former limits coerced acts to those that are not done intentionally.²⁶¹

As for the section comparable to Strawson's category 2b or altered psychology, Wallace addresses them as a separate group, which he develops in chapter 6 "Accountability and Exemptions" (following Watson's definition of exemption): those actions that are intentional, but done by someone who cannot be held responsible warrant a separate category of exemption. Cases where someone acts intentionally but is not blamed form a group apart from excuses according to Wallace. He writes that in cases:

in which the agent s intentionally does something of a certain kind (say, x-rather-than-y), and yet, [...] we would allegedly not hold s responsible for doing x-rather-than-y. The best way to make sense of this description, it seems to me, is to suppose that s's condition at the time of the action is such that it would not be appropriate to treat s as a morally accountable agent at all. [...] For only when an exempting condition is present will an agent who has done something intentionally be morally responsible for nothing that he has done (however carefully described). (Wallace 1994, 146-147.)

Within this exempted group of agents he includes, children, as Strawson did, as well as addicts and psychopaths, among others (ibid. 166-180; see also chapter 4.1.2).

²⁶¹ Frankfurt's theory of responsibility is not looked at here. It contains the idea that the agent identifies with her own actions and the idea that the agent is not responsible for actions that do not have alternatives (see McKenna 2009).

Whereas Watson and Wallace elected to use the word “excuse” where blame is modified to a certain degree but not eliminated completely in Strawson’s system, Holly M. Smith’s definition of “excuses” in *Encyclopedia of Ethics* (Becker & Becker (eds.) 2001) provides a general description of the excuse as a philosophical concept. According to Smith “An *excuse* is a condition pertaining to an agent that precludes his or her blameworthiness for wrongful action”, and in her case the word signifies complete elimination of blame for that action (Smith 2001, 506). It shares the distinction from justifications with Austin (1956/7) and additionally distinguishes the word “excuse” from “mitigating circumstances”. According to Smith, a *mitigating circumstance* “reduces but does not entirely eliminate blameworthiness”. (Smith 2001, 506.) Here the definition of excuse differs from Watson’s/Wallace’s version, which is the definition used further. Conversely, Smith describes the complete elimination of blameworthiness in the case of her excuse rather as Watson and Wallace talk about *exemptions*.

She states that the most interesting philosophical question associated with the topic of the excuse is “why [do] some conditions excuse an agent while others do not” (Smith 2001, 547). The theories addressing this question, of which Smith gives an example, include Aristotle’s theory of action (see chapter 4.1.1). David Hume, whose theory of responsibility is also very important, has a particular concept of character. In connection with Hume’s character, Smith refers to the “Hume-Brandt proposal” that an agent must have an intact character in order to be properly responsible for her acts. If the character is somehow incomplete or lacking, the agent simply has an excuse. This view, according to Smith (in addition to many of Hume’s commentators,²⁶²) finds the Hume-Brandt thesis and its variants problematic (Smith 2001).

Smith’s example for this problem involves a garbage collector who dutifully dumps a baby in with the garbage and argues that if the Hume-Brandt thesis²⁶³ is true, then the garbage man is not blameworthy in that the act did not follow from his character. Ignorance of the baby is unrelated to the character of the garbage man, and he would thus have an excuse (Smith 2001, 507). One could counter the argument, in that, were this the case, wouldn’t negligence leading to the dumping of the baby show that the garbage collector’s character inevitably had a rather serious flaw? For that matter, if he did receive an excuse (accidents will happen!), it might not be so strong that it would overrule any legal decisions, for example.

Smith mentions lists of excuses, usually including “*ignorance, immaturity, insanity, automatism, duress, necessity, coercion, compulsion, mental subnormality, mistake and accident.*” (Smith 2001, 507.) For the rest

²⁶² The commentator’s criticizing Hume’s character include Russell (1995) and Kinnaman (2005).

²⁶³ “Brandt assumes that most or all traits of character will be motivational states.” (Smith 2001, 507.)

of this dissertation, Watson's and Wallace's version is followed, rather than Holly Smith's, in that when "excuse" is mentioned it will be in the sense Smith uses "mitigating circumstance" and "exemption" as she would use the word "excuse". (See Smith 2001, 506-7.)

What have been discussed in this section so far were essentially four modes of speaking about excuses:

- Austin's "*ramiculated branch*" of philosophy

Since Austin's way of using excuse was linked to the discipline of philosophy of language/linguistics, a more precise concept was sought.

- Strawson's *special considerations* and *pleas*

Strawson used the terms special considerations and pleas to describe excuses. The type-1 pleas were such that, had the circumstances been normal, the agency of the person would not come into question. However, as the circumstances were out of the ordinary, or the agent was acting out of ignorance or under coercion, he or she is excused from responsibility. In the case of the second category, or the type-2 pleas, it is evident from the outset that the agent is not qualified for evaluation under the same moral agency criteria as any adult with normal mental capacities would be. Therefore, the cases which fall into this category are such that, the moral status of the agent renders any reactive sentiment completely eliminated considering these facts.

- Watson's and Wallace's *excuses* and *exemptions*

The distinction between the partial excuses and the complete exemptions were made initially in Watson's "Responsibility and the Limits of Evil" (1987) and has since been picked up, by R. Jay Wallace, among others. Essentially these terms are very close in meaning, so whichever set of concepts is selected is ultimately a matter of taste. Of the alternatives, Watson's and Wallace's version is used here because of its relative simplicity and the ability to recognize the terms as close to each other.²⁶⁴ Strawson's original definitions have often been criticized as being difficult and ambiguous, and the same criticism could be leveled at the distinctions between type-1 / type-2 pleas. Just the fact that the terms require explanation more than the others makes the original terms difficult to prefer.

²⁶⁴ E.g., John Doris in his *Lack of Character* (2002) also follows Watson's and Wallace's definition: "Responsibility attributions may be abjured in cases of deliberate behavior where the actor is in *excusing* or *exempting* condition." (Doris 2002,129.)

- Holly Smith's *mitigating circumstances* and *excuses*

Holly Smith described phenomena related to Strawson's pleas using the terms *mitigating circumstances* as well as *excuses*. In her case, *excuse* is used in the variety of instances where it is considered that the type of agency (or character) calls for total suspension of the reactive attitudes. In this sense, Smith's mitigating circumstances have the same meaning as Strawson's type 1-pleas. The term *excuse* corresponds with Strawson's type-2 pleas according to Smith's definitions.

As for Smith's selected concepts there is one additional reason why they are not used here, and that is the reservation on the word *mitigation*. Mitigation is referred to as the mechanism or the function of the character and excuses in the rest of the dissertation, meaning the way that the concept discussed can lessen the effect of the emotions related to wrong-doing and harm. Blame, as discussed elsewhere, is regarded as the distinctive element of moral responsibility, and one that is directly related to the latter. Thus, if an excuse is valid, blame is mitigated and furthermore the agent is considered responsible for the act as he or she would be otherwise (i.e., without the excuse).

	<i>Limited modification of reactive attitudes</i>	<i>Complete suspension of reactive attitudes</i>
Strawson	"Type-1" pleas or special considerations "expected to modify or mollify" the reactive attitudes. (FR, 7).	"Type-2" pleas or special considerations of an "agent [...] psychologically abnormal – or [...] morally underdeveloped", which are "expected to modify or mollify [reactive attitudes] or remove [them] altogether" (FR, 9).
Smith	mitigating circumstances	excuses
Watson, Wallace	excuses	exemptions

Table 1. Three definitions of moderate and complete modifications of reactive attitudes.

While Watson's and Wallace's definition is referred to henceforth as a rule-of-thumb, the concept of the excuse is not taken to be completely problematic.

Criticism of excuses

In an instance of questioning the current tendency to utilize the excuse, Erin Kelly in her article “What is an Excuse?” (2013) argues that being blameworthy is not always a required condition for wrong-doing from a Scanlonian position (see Kelly 2013, 246).²⁶⁵ Kelly challenges Strawson and Wallace, claiming that “persons might be excused [...] from being suitable subjects of blame, despite their violating the requirements of morality and acting wrongly.” (Ibid. 249.) Kelly emphasizes the cases that fall into the groups that are excused but that could be blamed for wrong-doing nevertheless. Kelly’s examination describes the “blind spots” of the Strawsonian moral theories that, when amended, could overcome some of their obvious deficiencies.

She discusses moral capability (supposedly in the vein of Holly Smith): “A morally capable person is capable of understanding morality’s requirements and so of regulating her behavior. Not only can she recognize that other persons are sentient, concerned about their future, have meaningful relationships, et cetera, but she also has a sensibility that moves her to care about these facts.” (Ibid. 250-251, see also Kelly and McPherson 2010.)

By recounting the discussion on Strawson’s and Wallace’s distinction between excuses and exemptions, Kelly alludes to Wallace’s statement that an “agent is exempted from accountability because that person lacks in a broader way, what Wallace refers to as “powers of reflective self-control”. These are the agent’s powers, generally speaking, to understand moral reasons and to thus regulate her behavior. According to Kelly: “Appealing to this model of defective moral agency does not do justice to the full range of cases [that Kelly has described at that point]”. (Kelly 2013, 254-255; see Wallace 1994, p. 157.)

Kelly proposes that “a reasonable assessment of when blameworthiness is mitigated or averted cannot depend on fine discriminations about capacity [required for moral capability],” arguing that it is too crude a scale to measure various defects of conduct as well as psychological pathologies. Instead, she claims, we usually follow our expectations of people being motivated. Excuses work by being able to intervene in these expectations of

²⁶⁵ Erin I. Kelly 2013 has written a recent article called “What is an Excuse?” (2013) in the *Blame* compilation by Coates and Tognazzini (2013), which discusses the concept of excuse in Strawson’s and Wallace’s theories. This skepticism toward the strict delineation of who exactly qualifies for moral assessments through the instrument of excuses, was expressed first by Gary Watson, and is discussed in chapter 4.3.4. Further on, how the concept of excuse relates to the idea of character is looked at in a broader comparison in chapter 6.

the way that the agent is treated separately from his or her intentions related to the act (Kelly 2013, 255-256).²⁶⁶

Kelly follows Scanlon by grouping wrong-doers' morally criticisable, retributive and non-retributive attitudes together as blaming responses, thus making it a broad definition of blame. She also notes standing to blame, by considering blameworthiness less positional to it. The case of child "soldiers" is given as an example, as there the agents appear to be less blameworthy than their adult counterparts.²⁶⁷ (Ibid. 246.)

Comparison between criminal wrong-doing, which contains evaluation of character and the circumstances in the actual sentencing process lead Kelly to advocate the use of a dedicated set of norms that can be consulted in determining "the moral quality of a person's action" and whether the agent is blameworthy. (Ibid. 249.)

She continues:

In challenging Strawson and Wallace's picture, I am drawing on the intuition that persons might be excused, in whole or in part, from being suitable subjects of blame, despite their violating the requirements of morality and acting wrongly. I will now take up some moral reasons we might have to maintain that an agent who is excused from blame can nevertheless have acted wrongly – the judgment Strawson and Wallace resist. (Kelly 2013, 249.)

The article advances through Strawson's excuses and pauses to consider the type-2 excusing conditions (see ch. 3.1.3) as "a range of cases in which [...] we draw a clear distinction between our moral appraisal and our varying attitudes toward the agent" (Kelly 2013, 248). Kelly claims that despite the conclusion arrived at regarding the agent based on moral appraisal of the action, we *are* able to hold them responsible by considering them blameworthy. In other words, where the agent is ordinarily exempted in the Strawson/Wallace model, there is actually a "scalar nature of excuses", which acts at various levels of mitigation.²⁶⁸ (Ibid. 249.) This seems intuitive and natural to accept; for example, if a 5-year old breaks a set of dishes for the fun of it, the parent can get upset and feel resentment. The latter does understand the role of the child and can act appropriately to discourage further misbehavior without the blame being as personal as it would had an

²⁶⁶ In terms of Bernard Williams's elements of responsibility (see ch. 4.3.3), what Kelly is referring to could be conceived as belonging to the state element.

²⁶⁷ It is unknown however whether Kelly considers the case of the child soldiers as standing to blame or of agency. If the child soldiers were facing incarceration by an adult regime, then it would be so, but merely comparing adult soldiers to child would seem more as a question of agency and the properties of moral capabilities, (which Kelly describes at length further on in the article.) (Kelly 2013, 246.)

²⁶⁸ Of the reactive attitudes, compassion is most closely associated with mitigating blame according to Kelly (Kelly 2013, 245, n1).

adult done the same. In the parent's case, it would be true as Kelly states: "A morally capable person is capable of understanding morality's requirements and so of regulating her behavior." (Ibid. 250.)

Kelly claims that the excuses are indicative of the agent's limitations in terms of moral capacity. These are to be considered when determining said agent's blameworthiness. Wallace's way of exempting agents who lack the "powers of reflective self-control" is unable to address the full spectrum of cases in which moral judgments are made, such as responsibility assessments (Kelly 2013, 255). This point is fully endorsed here, with concerns about Wallace's use of the encompassing explanation of "lack of self-control".²⁶⁹ Kelly uses the concept "incapacity thesis" to describe the demarcated classes of intentional but blameworthy action in Strawson's and Wallace's theories. Instances relevant to the incapacity thesis contain acts that can be judged bad, but are done by agents who should not be blamed for them. (Kelly 2013, 254.)

Kelly proposes that a reasonable assessment of when blameworthiness is mitigated or averted cannot depend on fine discriminations of capacity, proposing instead that "excuses function by undermining what normally are reasonable expectations about how a person should be motivated". Following Kelly's model would mean that excuses and blameworthiness (not blame, that is) would be understood as a mutually exclusive pair of concepts (ibid. 256-259).

Kelly concludes by referring to the difficulty that we are not required to "step in the shoes" of the wrong-doer in order to understand that viewpoint's condition fully, which would be the requirement for forgiveness and other types of understanding that would be fair in that situation. (Ibid. 262.) Overall, Kelly's article contains intuitions that are applauded here, but the new concepts (such as the incapacity thesis) that she offers do not seem very practical, at least compared with the definitions of excuses and exemptions dealt with above. Kelly's pointed criticism seems directed at the right areas of the Strawsonian theories. At this point it would be a mistake to claim that the concepts of character and excuse share a dichotomous relationship. The next section briefly describes two examples in which both character and excuses are relevant for moral responsibility.

²⁶⁹ See the "Criticism"-section in ch. 4.1.3.

6.3 EXAMPLES

Examples show that both sets of concepts, those based on character and those relying on excuses can be used to describe the same sorts of situations dealing with agency, especially those in which the agency is somehow compromised. A comparison between the theories as applied to these examples is made here, children first, and then the psychopath. There are of course only circumstantial similarities between these examples, and most authors are careful to keep them separated. They are looked at here simply because they are the most intuitive and obvious examples of conditions denying an individual full moral agency. The most obvious example, the moral status of children, can be described differently from the other, but with the same conclusion that the child is less likely to be considered blameworthy.

In the case of character, as with Aristotle's theory of responsibility (especially *NE* III.5) the character of the child is deemed undeveloped, and therefore lacking a capacity for moral decisions (as was the case with Irwin's complex theory, discussed in chapter 5.1.1), so that the child cannot be held responsible as an adult would in the same situation. From the viewpoint of a theory relying on the concept of the excuse, the child would be excused in the same way for whatever reason thought to be relevant that establishes (or disproves) the basic status as a responsible moral agent. The difference between the two is the emphasis on the historical phenomena in the case of character, as the subject was discussed in Fischer and Ravizza (1998, see chapter 4.3.3). Another group of agents for whom the moral agency of the person is fundamentally questioned from Strawson (2008/[1962]) onwards is psychopaths.²⁷⁰ Finally, one of the most recent sources that is of interest along with Watson's paper (2011), is Victoria McGeer's dialogue starting with "Lessons from Autism" in Sinnott-Armstrong's book *Moral Psychology vol. 3. The Neuro-science of Morality* (2008). McGeer discusses the relation between empathy and its lack in moral responsibility.

Children

Examples of children, as Eshleman writes, relate to the situation wherein the child is initially exempted from reactive attitudes and thus does not qualify for responsibility-assessments, but increasingly does through development gained by age. (Eshleman 2014.) The following examples detail a couple of groups that are nearly universally exempted from moral responsibility.

²⁷⁰ These other examples have been discussed in chapters 3.1.3 & 4.1.2. See also ch. 4.3.1, where early work by Fingarette (1966) discusses the psychopath based on an even earlier example supplied by Kant.

The groups with exemptions are often associated with corresponding philosophical examples. One of these examples deals with the underdeveloped position of children. Children are the natural common group that both Strawson and Aristotle exclude from normal objects of responsibility assessment. The children example is discussed in *Nicomachean Ethics* III.5, in which children and animals can act voluntarily but not deliberately. The acts children commit can be sudden and voluntary but, according to Aristotle, they are not *prohairesitically* chosen, and thus do not require a capacity for deliberation. (*NE* III. 5, 1111b8-10, see also ch. 5.1.2.)

As discussed above, in Aristotle's *Nicomachean Ethics*, the development and habituation of the virtues is discussed in a multitude of ways. The basic idea is that children, being morally underdeveloped, lack the requisite virtues to qualify them as full-fledged moral agents. While growing up, the young acquire the necessary virtues of character that allow them to be viewed as adults by the moral community. Aristotle holds that being blamed equates to being morally responsible (being blameworthy equates to the fairness of being held responsible), and while he discusses blame and praise together, the latter being seldom discussed in modern theories, there may be formulations of inverted thought-experiments in which the exempted status of the agent affects the response. Perhaps in some cases the inverted example can have inverse effects; for example, if a person who is responsible for a great deed and is praised for it is discovered later to have acted fraudulently this can negate that praise; a child might receive even greater adulation, since it was obvious that no one could expect the deed from him or her.

Broadie summarizes the inclusive attitude taken by Aristotle in *Ethics with Aristotle* (1991):

When Aristotle says that praise and censure apply to actions which issue from virtue and vice (developed conditions of the character), he is not, I think, forgetting the voluntary actions of the immature. Rather, the statement is about the terms in which we praise and fault. [...] Aristotle's theory of moral training assumes that under normal conditions, and with unspoiled agents, communications of this kind tend to be effective; he has a natural formal bent towards human excellence and happiness, but initially depends on others to supply a content. (Broadie 1991, 127-128.)

Despite this, children are clearly exempted from moral responsibility ascriptions as a result of Aristotle's tendency to view them as categorically separate from humanity. Broadie writes: "In Aristotle's ontology, a child or childlike member of the species is not a complete human substance; it could no more ethically exist apart from guardian or guide than a foetus physically could from the mother-animal." (See Broadie 1991, 64.)

In *FR* Strawson considers children very briefly, but he does note them as significant among the groups of agents listed under the type-2 special

conditions that serve to completely eliminate the reactive attitudes associated with the action taken by these persons. In connection with the second subgroup of type-2 conditions, Strawson says that the reactive sentiments (including blame) of both children and the mentally deranged are modified or mollified.²⁷¹ Garrath Williams also notes that in terms of Strawson's reactive attitudes "our relations with children occupy a crucial in-between place" (G. Williams 2010).

Strawson discusses the undeveloped foundation of the morality of young children in the following way:

Thus parents and others concerned with the care and upbringing of young children cannot have to their charges either kinds of attitude in a pure or unqualified form. They are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet truly capable of either. The treatment of such creatures must therefore represent a kind of compromise, constantly shifting in one direction, between objectivity of attitude and developed human attitudes. [...] The punishment of a child is both like and unlike the punishment of an adult. (FR, 20-21).

Emphasizing his conclusion, Strawson gives an example concerning upbringing: children are increasingly capable with age in the sense of being full members of the moral community, but until they are that, they cannot be fully considered as such. This leads to a suspension of the reactive attitudes as described above.

As for the commentators on Strawson, in "Responsibility" (1966) Herbert Fingarette discusses children prominently, as he seeks to evaluate our understanding of responsibility using the example of a five-year-old child causing a fire. The example helps to distinguish the causal responsibility of the act of playing with matches with the actual event of the fire from blameworthiness associated with the act. Although Fingarette is not specific about the formation of the judgments related to this situation, he does imply that we do not consider five-year-old children blameworthy. More recently, Pritchard's *Becoming Morally Responsible* (1991) is devoted to exploring the development of personality toward morally responsible agency. This rare example of informing ethics with developmental psychology is interesting in terms of children and the boundaries of the concept. Additionally, as discussed, Wallace's discussion of the children in *Responsibility and the Moral Sentiments* (1994) is found in the sixth chapter, where they are treated as the first exempted group of agents (Wallace 1994, 166-167). Wallace concludes that not having developed reflective self-control, children are exempted from blameworthiness.

²⁷¹ "The agent was himself; but he is warped or deranged, neurotic or just a child" (FR, 9).

Psychopaths

While in Aristotle's case anything said about psychopaths is unavoidably speculation, we may pursue some interesting angles by applying Aristotle's theory to the problem. There is a current, active debate on the moral responsibility (and criminal responsibility²⁷²) of psychopaths, which is epitomized by a paradoxical stance toward them. Perhaps the most relevant outlook for the present purposes, is that by Gary Watson who in "The Trouble with Psychopaths" (2011) describes the problem that philosophy has with the psychopath. The missing component required for the moral analysis of the acts of a person afflicted with this condition, is identical with that whose absence causes the aberrant behavior.²⁷³ The component in question is essentially *empathy* (see Watson 2011, 323-324), according to which psychopaths should be treated as an external group distinct from normal moral agents.

As Bobzien (2014) relates, Aristotle has a distinction of agency that could help to distinguish psychopathic agents who can be held blameworthy from those who cannot. Bobzien lists types of agency in her article "Choice and Moral Responsibility" (2014) 1. "The Virtuous Agent" 2. "The Strong-Willed Agent" 3. "The Weak-Willed Agent" 4. "The Vicious Agent" and 5. "Agents with Not Yet Fully Developed Character". Where the theory of Aristotle could recognize cases as blameworthy, psychopaths would belong to the two last groups of agents. An agent acting in a vicious manner who is still an adult (and not a child or childlike as in the previous example) would be blameworthy in Aristotle's theoretical context. (Bobzien 2014, 101-105.)

What is interesting is that Aristotle conceives an adult who has not gained sufficient character traits as a full member of moral community, yet he still holds such an individual morally responsible. Bobzien claims that given these conditions Aristotle's theory reserves the option to hold an adult but incomplete agent morally responsible: there are four conditions for this: 1. The moral dispositions known as virtues have not formed yet to form a full character. 2. The agent is capable of practical reason, i.e., to pick the golden mean. 3. He/she has appetitive wanting (*boulesis*), which is distinct from choosing deliberately. 4. He/she acts voluntarily, having knowledge of alternatives; if the chosen act is based on the vices, then these conditions are applicable. (Bobzien 2014, 104-105.) Obviously these points are not enough for a complete picture of "Aristotelian psychopathy", but this topic might be interesting enough to warrant further research.

Meanwhile, referring to the psychopath, problems arise with the Strawsonian default model of responsibility with those agents who recognize the criteria of "qualifying as a responsible person" even when they could in

²⁷² See Brink & Nelkin (2013) as well as Godman & Jefferson (2014).

²⁷³ "The trouble with psychopaths is that they at once meet the malice test but fail the moral competence requirement." (Watson 2011, 309.)

fact have a radically different set of morals. (See G. Williams 2010, Watson 2011.) Strawson doesn't devote much thought to the examples he lists. He does however touch upon the subject of someone who is described as a "moral idiot" (*FR*, 13). More broadly, psychopathy has been a major theme in Watson's articles as evidenced by "Responsibility and the Limits of Evil" (1987) and "The Trouble with Psychopaths" (2011). On his account, the key challenge with philosophers discussing the position of the psychopath is that this aberrant behavior arises from deficient mental qualities – the same qualities also fundamental to the requirements of treating each other as moral agents. In Pritchard's book *Becoming Morally Responsible* (1991), chapter 3 is considered by a few influential writers as important. These commentators maintain that his book is very relevant in the discussion regarding the morality of the psychopath. In *Responsibility and the Moral Sentiments* (1994) Wallace equated psychopathy with a lack of reflective self-control. Likewise he wrote of children that by being morally undeveloped children lack the requirements for reflective self-control. According to Wallace "Such agents are for instance nonmembers of our moral community, insofar as they do not accept the moral requirements that we accept and hold people to, and they may be incorrigible in their immorality as well." (Wallace 1994, 178.)

A more recent commentary by Victoria McGeer includes psychopathy in a very interesting moral psychological assessment of the condition of autism and other forms of extraordinary mental states in her article "Lessons from Autism" (in Sinnott-Armstrong 2008). While both autists and psychopaths "seem to lack empathy in some sense of that word, and yet [of these two groups, only the] psychopathic individuals are well known for their lack of moral concern [...] What accounts for this difference?" By answering this question McGeer seeks to shed new light on the nature of moral agency. (McGeer 2008a, 228.)²⁷⁴

The two examples of children and the psychopath clearly point out the similarities between the two varieties of mitigating blame. Different reasons can lead to alterations in moral judgments. In practice, the outcome is the same, no matter which concept is used to explain blame mitigation. For instance, a child is pardoned from breaking things no matter whether it is just excused by his or her undeveloped moral agency, or a more intricate analysis of formative events in his or her character history. While the

²⁷⁴ Victoria McGeer's article (2008a) is doubly interesting as it is directly followed by a comment by one of the authors that McGeer herself is debating with, Jeannette Kennett, which in turn is followed by McGeer's response. According to Kennett, McGeer assumes that empathy is a necessary requirement for morality, and while it might be the most popular "route to morality", Kennett stresses that "it is not the only one" (Kennett 2008, 260). McGeer's reply restates her stance that "we are reasoning creatures with a certain range of affectively determined concerns. Take away either the affective component or the reasoning component and you take away our capacity for moral agency." (McGeer 2008b, 283.)

explication of this state of the agent's mind have been very brief here, they are considered as their own special, separate groups in the preceding chapter (7), which compares the theories in which such extraordinary agents may influence the moral responsibility assessments.

6.4 SUMMARY

As Ferdinand Schoeman hinted in his introduction to *Responsibility, Character and the Emotions* (1987) character and the excuses share a connection (see Schoeman 1987, 9). If Bernard Williams is correct in claiming that every theory of responsibility must include the four elements of responsibility, including blame as a response; and if the current debate on blame is considered germane then blame is a key element of responsibility relations. From this assumption – and it is considered reasonable to accept this assumption here ²⁷⁵ – it follows that every theory of responsibility contains means of discussing the appropriateness of blame in the cases of differences of quality between the agents. When this appropriateness is questioned, the initial reactive sentiments and resulting blame can be re-evaluated, based on which, for lack of a better term, the blame can be mitigated. This sort of *blame mitigation*, is thus present in one form or another in the theories of responsibility; either as conceived in terms of the broad descriptions and conditions related to character, or with the more specific and immediate excuses.

In general, *character* is something that has been discussed in connection with responsibility, but has received less attention since the last century. As the methods and accuracy of psychological scientific methods have advanced, the area covered by this concept has been both clarified and reduced. Nevertheless, philosophical theories of responsibility have continuously referred to character as one way of reducing, mitigating or even overriding blame. As other theories have been content to provide lists of excuses for this same function, a comparison of how these ideas fit together follows in the next chapter.

As an agent commits an act, responsibility can be defined as consisting of the response that the surrounding moral community gives to that act. Now if the agent is extraordinary, excuses that may or may not be based on character can apply. In the case of such an excuse, the blame is mitigated or cancelled out. In everyday interaction, there are no action logs, or lists of conduct to be checked. The views criticizing "the ledger view" of responsibility are followed in this sense. Character could be described as sort of time-molded stamp²⁷⁶ that can be applied to the moment of the action and direct the course of action. Thus the state of character of the person consists of dispositions and thought patterns that set the probabilities (see McIntyre 1990) that a particular action is chosen.²⁷⁷ While character is however never

²⁷⁵ See chapters 4.1.2; 4.2; 4.3.3 & 5.1.1.

²⁷⁶ The original Greek word *charaktēr*, which means the imprint on a coin (Homik 2015).

²⁷⁷ Jane McIntyre's article "Character: A Humean Account" (1990), published in *History of Philosophy Quarterly*, comprehensively discusses the peculiar way that Hume uses the concept of character in his work. She argues that it is possible to accept Hume's controversial idea that all

a sufficient cause for blame or exemption, it is a necessary element²⁷⁸ when moral responsibility is considered, just as a child's undeveloped character is also a reason for exemption.

Character, considered as a distinct concept, is a vaguely defined thing with a strongly subjective element. What a character is from the perspective of the agent and how it is interpreted from the viewpoint of the community is different, but the two conceptions can be brought closer together where more information is available. How this information is increased depends on the situation: the various parties could converse and get to know each other better, or some relevant background histories of the agent could be uncovered by searching documents, etc.

Two alternatives for mitigating blame in theories of responsibility were examined here: character and excuses. Neither of these was without problems: the idea of character, originating with Aristotle, expresses the effect of history, of previous events, and the various causes related to the act being evaluated. Suffice it to say, since accounting for every event prior to the wrong-doing is impossible, and without the intention of excusing every single act, some other criteria are needed to determine the blameworthiness of the agent. Additionally, it seems that the two different concepts are not identical as, for example, there are excuses that are not related to character (at least when the condition of being related to character is not enforced, like Strawson's type-1 excuses). This corresponds with the caution at the beginning of the comparison about not treating the two concepts as being in a dichotomy with each other.

Character, as stated above, implies that the actor has a choice, and that her character traits such as virtues can affect this choice. If the character of the agent is incomplete, as it would be in the case of a five-year-old child, or if its functioning is otherwise disturbed or suspended such that the acts are somehow done out of character, the agent is not blamed.

Excuses on the other hand do not imply that the act has alternatives, although it is possible that there are. Excuses imply that the act committed is wrong, but that for some reason the agent is not blameworthy. The excusing conditions, as described by Strawson, provide a handy instrument to circumvent the excessive historical plotting needed if moral responsibility is conceived in terms of character. If the agent, her moral status, or a trait of this status warrants an excuse, the agent is not blamed, or if she is, according

responsible action must originate from the character, if interpreted in a certain way. The article is very helpful in gathering Hume's fragmentary writings on the subject of character, and in selecting what is essential for moral responsibility. The article seems to be the formative work of the current interest in Hume's character. In a 2015 Hume Society conference panel, McIntyre made comments that followed up on issues in the 1990 article, and that have since attracted comment from other scholars. While regretfully not germane to the entire dissertation, some of these additions are noted here anyway, with the kind permission of the author.

²⁷⁸ Bernard Williams's state element (Williams 1993 & 1997).

to Wallace, it would not be fair. But without additional external description these excuses can appear arbitrary and baseless and as a result, as is the point of “the ledger-view” of responsibility, without the relevant cultural context it would be more of a metaphysical concept than actual practice. Both lead ultimately to the same conclusion that the agent is not blamed, or that the blame that would be directed toward the actor in normal circumstance would in this case be withheld.

Reasoned this way, character and excuses could be viewed as different means to the same goal, which is aggregated here as blame mitigation.²⁷⁹ The cross-referencing of these two “means of blame mitigation” might not be the most pressing activity in the current philosophy of moral responsibility, but some attempts have been made or are being made currently. Nevertheless, this exercise has shed light into some corners of the topic that have seldom been thoroughly scrutinized.²⁸⁰

In conclusion, setting up character and excuses against each other would be a false dichotomy. Instead, what could be useful would be to sample comparable examples of how different theories treat the subject matter in relation to the terms and the relation between these concepts explained by this comparison. The theories which are partial to character-based explanations would not automatically prevent excuse-based approaches. The following combinations of attitudes toward the concept are conceivable:

a) The first group embraces character, and rejects the excuses. This is a relatively rare option to choose between the theories, as it tends to lean toward the incompatibilist accounts.²⁸¹

²⁷⁹ Thanks to Sæde Hormio for influential criticism leading to the selection of blame mitigation as the point of contrast used here.

²⁸⁰ As mentioned in chapter 5, Irwin compared Hume’s theory of responsibility with that of Aristotle (Irwin 1980, 134-135). Hume’s stance on the concepts of excuses and character were viewed as a sort of compromise, both concepts of character and excuse co-existing within the limits of his theory of responsibility. Surprisingly few have followed suit, the supposed reason being the fear of being branded as an incompatibilist. As Watson’s 1987 examination of the concept showed, there is more to the idea of historical responsibility than is usually admitted. Even if this aspect is not decisive in responsibility assessments, it certainly has an effect on our reactive emotions. By following Bernard Williams’s elements of responsibility, the practice of blame mitigation could be restated in a more elegant form. This could in turn provide further instruments for the ever-problematic explication of responsibility. For now, however, thorough examination of Hume’s responsibility is saved for a future occasion.

²⁸¹ One example of these is the book *Against Moral Responsibility* (2011) by the hard determinist Bruce Waller. Waller argues that Strawson and his followers fail to distinguish between character faults and blame faults. One of the examples he uses is Watson’s Harris example as he argues for the weight of character concerns being above the concerns of blameworthiness in determining responsibility. (Waller 2011, 153-159.) This has to be considered, however, within the limits of Waller’s main argument that moral responsibility and its related practices are “utterly eliminated” (ibid. 7). One interesting term that Waller uses is *excuse-extensionism* (presumably after Smilansky (2000)): Waller

b) The other “either-or” option most frequent in the post-Strawson scene is that the excuses should be embraced and the character considerations rejected. Wallace is the most pre-eminent of these “excusers” as he rules out the character-related concerns as a “sort of deep responsibility for one’s actions” and being part of “the condition of autonomy” by stating that they should not be confused with being an actual element of moral responsibility. (Wallace 1994, 52-56).

The remaining two options are c) attempting to include both character and excuses as complementary means of blame mitigation. These are few, but Hume is the obvious theorist who should be examined further on this. How significant is this exactly? Going back to Hume might be useful as his conception of responsibility did have both character and excuses in its repertoire. Character acts as the essential basis of moral responsibility in his theory as well as excuses.²⁸²

The final group is d) rejecting both. This could be conceived through looking at other research that has been done. An example is offered here: Kelly (2013) criticizes excuses as insufficient to determine blameworthiness and moral agency as an extension. She does this without forming a new theory as such. She does however follow Scanlon in terms of blame, which needs to be taken into account, if one wants to stay in the Strawsonian context. (Kelly 2013, 246-247.)

thinks the moral responsibility abolitionist position may contain a fundamental error. According to this concept, there are valid compatibilist excuses; that much is given. The position however tends to lead to erroneous elimination of moral responsibility through at least initially well-meaning extension of the excuses. (Waller 2011, 212.) Daniel Dennett rebuts Waller’s criticisms in his review of *Against Moral Responsibility* (2011). He praises Waller’s scholarship on the subject, but claims that he misinterprets responsibility (in order to emphasize the current bad state of affairs within the U.S. system currently), and that by abolishing moral responsibility as a system he makes insufficient distinction between the original and the replacing concept of *take-charge responsibility*, which he is producing. According to Dennett, this is essentially the same concept as his moral responsibility. (Dennett 2011.)

By focusing on the retributive side of the concept at the expense of moral responsibility as *making moral judgments*, Waller’s stance is hard to accept as a basis for his argument, and as a result it is difficult to concede his conclusion of abolishing moral responsibility as such. Approached with different starting premises, the work might be more appealing.

²⁸² According to Jane McIntyre (1990) Hume considers character as the necessary basis for excuses. Paul Russell (1995) and Ted Kinnaman (2005) do however criticize him for this. Paul Russell argues (from a Strawsonian perspective) that Hume did in fact make an error in this case. (Russell 1995, 95-105.)

7 ANALYSIS

A clear explanation of how the concept of blame is central to accounts of both Strawsonian and Aristotelian responsibility needs further examination. In the previous chapter, the concepts of character and excuses were discussed, and it was concluded that both of these could function as forms of blame mitigation, or in other words that they could be the mechanisms for reassessing how the initial reactive attitudes related to action could be based on additional information. In this chapter, Bernard Williams's explication of the elements of responsibility (Williams 1993 & 1997; discussed in chapter 4.3.3) is brought back in the form of the framework of analysis, with the help of which it is examined how character and excuses could both signify the means to mitigate blame. These concepts, as parts of the two different accounts of moral responsibility discussed above, can be analyzed side-by-side with the help of these elements. The elements of responsibility are presented with the help of tables, in which the relevant concepts that appear in the theories of responsibility are compared with each other. The interaction of these different elements is described up front in terms of simple propositional logic in order to clarify the relations between the concepts. The three different theories of moral responsibility analyzed are: Aristotle's, Strawson's and Wallace's.

To reiterate, further detail of the "fabric" of responsibility can be adduced by utilizing Bernard Williams's distinction into four elements of responsibility. He argues that all theories that discuss responsibility stress one or more of the four following elements of responsibility: 1. cause, 2. intention, 3. state and 4. response.²⁸³ Williams claims that these four elements, common to all attempts to classify responsibility, can be distinguished in those theories to varying degrees. Cause refers to causal responsibility, while the difference "between *intention* and *state* is one between what an agent means to do [...] and what state of mind he is in when he means to do it and does it" (Williams 1997, 97). Because these are rarely dealt with identically between the different theories, ultimately there can be no universally acceptable definition or concept of responsibility, as so much depends on particular circumstances.²⁸⁴

Here Williams's elements are used as a framework for the thought experiment that whenever the appropriateness of blaming is questioned, these elements are tested: the first element is true if the agent is causally responsible (CR(sx)) for the act and false if not. The second is true if he or

²⁸³ Williams (1993), chapter 3 & 1997, 96.

²⁸⁴ "There is not, and there never could be, just one appropriate way of adjusting these elements to one another – as we might put it, just one correct conception of responsibility." (Williams 1993, 55.)

she intended ($I(s^{t(x)}x)$) to do the act and false if he or she did not. The third element of the state ($S(s^{t(x)})$) is problematic. If true, it implies that there is nothing out of ordinary about the state of mind of the agent. If false on the other hand, it means that there are some active considerations that affect the evaluation of the acts done by the agent. If all three of these elements of responsibility are true, the agent can be considered morally responsible ($MR(sx)$) for the act. Moral responsibility in turn finally implies that the response element ($B(sx)$) is true as the blame could be assigned to the actor because of the previous elements permitting it to be so. Of course, where blame is mitigated (as it is in most examples given here), this element is false.

The four elements of responsibility are used here in connection with a simple true/false condition of whether they matter for the excuse. We see below how the analysis is conducted: if the cause element is true, it means that s is causally responsible for doing x .²⁸⁵ In this case, the first condition is true in every one of the examples, because the interest is whether s is morally responsible, and whether the resulting blame is mitigated or not. If the cause element is false, the concerns about responsibility would refer to some kind of “scapegoat” situation or to events that have not happened; in other words, metaphysical guilt.²⁸⁶ Since being a scapegoat effectively eliminates the possibility of moral responsibility and since the latter cases are relevant only in terms of imagination and motivation, the nonexistent events are of less concern. As the Strawsonian approach emphasizes action, the former are not looked at here. In order to produce the comparison between the theories, there are really only two claims that describe the relations between the functions:

1. $CR(sx) \wedge I(s^{t(x)}x) \wedge S(s^{t(x)}) \rightarrow MR(sx)$

2. $MR(sx) \rightarrow B(sx)$, where:

- $MR(sx)$: s is morally responsible for x ;
- $CR(sx)$: s is causally responsible for x ;
- $I(s^{t(x)}x)$: S intended x (at the time of x or shortly before);
- $S(s^{t(x)})$: S 's state of mind is that of a morally capable adult human (at the time of x or shortly before) and
- $B(sx)$: S is blameworthy for x .

In normal cases, where the morally responsible agent is blameworthy, he or she is also considered to be morally responsible in the broader sense. As

²⁸⁵ As discussed in ch. 4.2 “causal or explanatory blame” was all but dismissed by Coates and Tognazzini (2014).

²⁸⁶ The concept of metaphysical guilt, where each condition but causal responsibility is true, has been considered at least by Larry May in his book *Sharing Responsibility* (1993), but, again, this is a topic for another occasion.

stated by Williams, this responsibility is an aggregate condition of different concerns: the agent is causally responsible for the act, he or she intended it and finally, the state of mind of the agent was in the fully operational condition of an adult human being (or of sufficient maturity dependent on the situation).

Just to describe a normal situation, where A does x to s and is causally and morally responsible and therefore blameworthy, you can think of a well-to-do gentleman thief who plans a heist in a jewelry shop and commits it just for fun and personal gain. As the thief breaks into the shop, Williams's cause element is true, because he breaks the window of the shop. The intention element is true, because of the conscious planning and the commitment to carry the crime through. The state element is also true because the hypothetical burglar is in good spirits and has no severe mental problems whatsoever (think *Ocean's Eleven*). He's not even in need of the money. The previous conditions all being true, since there is no reason for blame mitigation, the condition for the response element is true as well. As all four elements are true, the wrong-doing is clear, and no alleviating concerns are present, the blame is directed at the perpetrator as a result. In such a situation, moral responsibility blameworthiness would not be mitigated. The agent is considered a morally responsible adult human being. In terms of Williams' elements, a normal responsibility assessment would be:

1. $CR(sx) : \text{true} \wedge I(s^{t(x)}x) : \text{true} \wedge S(s^{t(x)}) : \text{true} \rightarrow MR(sx) : \text{true}$
2. $MR(sx) : \text{true} \rightarrow B(sx) : \text{true}$

In contrast, the most interesting of possible alternatives discussed here is the one in which the agent's state of mind is not that of a morally capable adult human. Here the state element would be false. In terms of the elements of responsibility, this example would appear as:

1. $CR(sx) : \text{true} \wedge I(s^{t(x)}x) : \text{true} \wedge S(s^{t(x)}) : \text{false} \rightarrow MR(sx) : \text{false}$
2. $MR(sx) : \text{false} \rightarrow B(sx) : \text{false}$

In this second situation, the state of mind of the agent can have an effect on the responsibility assessment for the act regardless of the relationship between the agent and the act.²⁸⁷ This implies that there are conditions related only to the agent or the status of the agent that can affect the final responsibility assessment. For simplicity's sake, further references to the above functions are represented by tables showing the relevant outcomes by just referring to the truth value of the corresponding element of responsibility.

²⁸⁷ Kelly (2011) might argue against this.

7.1 ARISTOTLE'S CHARACTER-BASED BLAME MITIGATION

Williams's elements within Aristotle's theory appear in terms of the following topics:

Williams	Aristotle
Elements of responsibility (Williams 1993, 55 & 1997):	Content of said element:
Cause	Efficient cause (<i>aitia</i> / <i>aition</i>)
Intention	Action (X) is voluntary, meaning it is a1) done in knowledge a2) not coerced. ²⁸⁸
State	The agent has a sufficiently formed character, i.e., is an adult capable of decision (<i>prohairesis</i>) (NE III.5.)
Response	Praise or blame (NE III.1. & EE II.6)

Table 2. Bernard Williams' elements of responsibility in Aristotle's theory of responsibility.

In table 2, *aitia*, which is mentioned as a term employed by Aristotle to describe notions related to responsibility, refers to *efficient cause*. This is in contrast with Aristotle's other causes, which include the material, formal, and final causes, and which can be additional extensions of the direct translation. As discussed in chapter 5.1.2, the teleological cause may have a distinct effect on the action and the responsibility relations related to the act. Williams's cause element is represented best by Aristotle's idea of the efficient cause, as it exemplifies the situations in which someone (s) is directly causally responsible for an act (x) or an event. As for the intention element, the Aristotelian excuses (chapter 6.1) describe intentionality fairly well. Intentional acts are done in knowledge and uncoerced. How the state is discussed in NE III 1-5 is less explicit. However, seeking the elements shows quite clearly that NE III.5 deals with cases where the state of mind of the actor is not up to par in moral agency. Finally, regarding the response element, Aristotle's theory of moral responsibility offers its analogue at the beginning of book III.1. Book II.6 of the *Eudemian Ethics* is also quite clear

²⁸⁸ See, for example, Irwin's "simple theory" (1980) & Roberts (1989).

on the stance, which puts praising and blaming in direct connection with responsible action.

The alternatives, where blame is mitigated in connection with the relevant elements of responsibility are shown in table 3 below. If the element is marked as “true”, it means that in these cases the corresponding element is effective in that situation. Correspondingly, if the element of responsibility is marked “false”, this means that as a requirement for the agent to be morally responsible that element is incomplete or untrue, preventing the agent in question to be held blameworthy for act X.

Aristotle	Williams
Action-based excuses:	Elements of responsibility:
a) Involuntariness	1. Cause: true
a1) Coercion	2. Intention: false
a2) Ignorance	3. State: true
	4. Response (praise/blame): false (as in not blameworthy, not morally responsible)
Character-based excuses: b) the agent’s character belongs in one of the following groups: animals, children	1. Cause: true
	2. Intention: true or false
	3. State: false (due to lack of decision capabilities)
	4. Response (praise/blame): false

Table 3. *Aristotle’s action-based and character-based excuses compared to Williams’s elements.*

Thus, Aristotle’s theory of moral responsibility discussed in terms of Williams’ elements would imply that the interesting extraordinary cases of excused action would be those that fail the condition of intention and those that fail the state element. In the former case, the agent cannot be held blameworthy because that act x was done either out of coercion or ignorance.

In the latter, the agent fails the blameworthiness qualification because of his or her character. In other words, by being a morally underdeveloped child, mentally ill or indeed by not being a member of the humankind in the first place, as with animals in Aristotle’s theory. As he thought children were part of a different substance altogether than adult human beings, a different method of measure is needed in the various age groups. In his theory, it is not possible to evaluate the moral responsibility of a child without the blameworthiness of the guardians of that child. Furthermore, the requirements for practical reason, *prohairesis*-type choosing and a set of

virtues are required for adulthood. Animals and people with lacking decision-making capabilities from any reason are considered separate from the group of blameworthy individuals (see chapter 6.1).

7.2 STRAWSON'S AND WALLACE'S EXCUSE-BASED BLAME MITIGATION

This section shows how Williams's elements appear within the Strawsonian theories in *FR* and the way Wallace (1994) expands Strawson's categories is demonstrated and analyzed. In table 4, the elements of responsibility are shown as they appear in Strawson's context:

Williams	Strawson
Elements of responsibility (Williams 1993, 55 & 1997):	Content of said element:
Cause	Causal responsibility
Intention	Action of the agent is not accidental or caused by circumstances (<i>FR</i> , 7-8.)
State	The agent is considered as part of the moral community i.e., held to be a fully capable adult (<i>FR</i> , 9-11.)
Response	Reactive sentiments (<i>FR</i> , 4-7.)

Table 4. Bernard Williams's elements of responsibility in Strawson's theory of responsibility.

Causal responsibility is, as Eshleman (2001/[2014]) noted, not as significant in the modern theories of responsibility such as Strawson's. Nevertheless, the causally-initiated event is considered to be the prerequisite for proper responsibility assessment to take place. Intention however is inherent in Strawson's examples by way of negative definition. His pleas or special circumstances (as described in chapter 3.1.3) assume that the action is caused by an unintended accident or unusual circumstance, and thus the agent is not as blameworthy as he or she would be without these interfering factors. It seems that the cases that are described in the first category of Strawson's excuses "unintended consequence and coerced acts", are those where the cause and the state elements are true, but the intention is not, and thus the response or the blame might be mitigated.

These excuses are also devoid of intention, but a diminished state of mind might also be a factor. In any case, the blame is mitigated as a result of temporarily suspended reactive attitudes. Strawson's second sub-group of excusing conditions is called here the group of "Abnormal psychology". Williams's intention element might be true or false in these cases, but Strawson considers the state element as the more significant in viewing the agent in a different "light" because of the observed state of the agent as that condition effectively eliminates blameworthiness entirely. Strawson's special circumstances in relation to Williams's categories are described in table 5 below:

Strawson	Williams
Excusing conditions (X):	Elements of responsibility (in case of X):
1. Unintended or coerced acts (agent normal, less blameworthy)	1. Cause: true 2. Intention: false 3. State: true 4. Response (blame): false
2a. Abnormal circumstances (reactive attitudes suspended)	1. Cause: true 2. Intention: false 3. State: true or false 4. Response (blame): false
2b. Abnormal psychology (reactive attitudes modified or eliminated)	1. Cause: true 2. Intention: true or false 3. State: false 4. Response (blame): false

Table 5. *Strawson's excusing conditions compared to Williams's elements.*

In a similar cross-referencing of Williams's elements of responsibility with Aristotle's theory, Strawson's theory most clearly contains the analogues of the type-1 excuses where something is lacking in terms of intention. Equally clear is the second sub-group of "abnormal" or extraordinary psychology²⁸⁹. In these cases, the state element has a negating effect on the reactive instances of resentment or gratitude that is derived from act X.

Being the only category devoted to completely external influence, the first sub-group of Strawson's type-2 excuses (2a) is difficult to compare with the other excuses. This one is characterized by abnormal circumstances. While it

²⁸⁹ Since "abnormal" for Strawson includes children, the word might be too strong at least when considered from outside. This problem of Strawson's theory is discussed by Kelly (2013). among others.

may be true that the agent's moral capacities cannot be questioned initially, the variety of ways in which the circumstances can have an effect on the mental capabilities of the agent suggest the possibility that both the intention and state of mind of the agent are compromised in these situations. Nevertheless, the end result in terms of response is the same in each case: regardless of the type of excuse, the agent is less appropriate as a recipient of praise or blame.

Wallace's categories do not fall as easily into the clear divisions of Williams's conditions as those of Strawson, but one thing shared between the types of excuses from a) to d) is that the condition of the cause being true and the condition of intention being false. The only exceptions to this are excuses a) and b), which were not sufficient reasons for blame mitigation by themselves, because the possibility of negligence by the agent leading to the act x. Thus, if an inadvertence, mistake, accident or unintended bodily movement is the result of the agent's own negligence, the response is true, i.e., blame is not mitigated. In excuses c) and d) the state element could be either true or false, but the end result is the same. The lack of intention is enough to question the appropriateness of blame, which is mitigated regardless the state of mind of the agent. Table 6 is devoted to describing the elements of responsibility in Wallace's context:

Wallace	Williams
Types of excuse:	Elements of responsibility:
a) Inadvertance, mistake or accident	1. Cause: true 2. Intention: false 3. State: true
b) Unintended bodily movement	4. Response (blame): false / true if negligent
c) Physical constraint	1. Cause: true 2. Intention: false 3. State: true or false (inconsequential ²⁹⁰)
d) Coercion, necessity and duress	4. Response (blame): false
Exempting conditions: e) The agent is a member of one of the following groups: children, the mentally ill, addicts, the hypnotized, psychopaths or the heavily stressed	1. Cause: true 2. Intention: true or false 3. State: false 4. Response (blame): false

Table 6. *Wallace's excuses and exemptions compared to Williams's elements.*

The differences between Strawson's and Wallace's categories are slight: Wallace's a) and d) are essentially the same as Strawson's category 1. Wallace's new categories b) "unintended bodily movement" and c) "physical constraint" could fall into either of Strawson's groups, depending on the reason why the excusing condition is present.

Wallace's elaboration of the Strawsonian excuses produces fairly straightforward analogues of Strawson's examples, with a couple of additional clarifications: in the case of those basic situations that Strawson would have thought of as examples of special conditions or pleas from the first group, Wallace argues that if, in these cases, the cause of the unintended or inadvertent act stems from negligence, the excuse might be invalid (see Wallace 1994, 142). Otherwise, further explication of the categories of excuses yields little alteration to Strawson's counterparts. The case of inadvertent negligence is interesting, as the elements are not directly comparable with the general statements that were drafted at the beginning of

²⁹⁰ Although in the case of coercion Saba Bazargan (2014), as a consequentialist, would argue that there are degrees of coercion that might result in *shared* liability between the coercer and a wrong-doing coercee. In these cases, the wrong-doer might be eligible to compensate some of the harm caused. Generally, according to Bazargan, the majority of the accountability falls on the shoulders of the initial coercer. (See Bazargan 2014, 12.)

the analysis. There the intention element toward the act in question would suggest that, as the negligent action causes unintended events, blameworthiness should be questioned. However, this again is a situation requiring a longer observation period, to the point where a crucial detail or an act is intentionally omitted or taken, leading to the negligent behavior. A prime example is drinking and driving, where the conscious, uncoerced decision to start drinking in the past would still qualify the condition of intentionality for moral responsibility for the later events, where such deliberation no longer occurs.

Concerning the examples dealing with uncertain cases subject to the effects of circumstances, it can similarly be asked whether the state element is or is not relevant. As always, the unintentional nature of these cases related to unusual circumstances is sufficient by itself to allow doubts about blameworthiness. As a result, the exact nature of the agent's state of mind ultimately remains secondary, or even inconsequential by definition.

An interesting example of these cases with a formidable effect on the circumstances occurs where the state element of the agent actively works against a failed intention element. One such example could be an unafraid fire-fighter who is seen running away from a burning house because he or she had seen someone in danger on the other side of the road. Alternatively, a well-known billionaire benefactor who is recognized as a magnanimous philanthrope fails to contribute toward an essential cause because of an error in bank statement service. In these cases, the circumstance forces an otherwise intentional, morally praiseworthy act to appear as something deserving blame.

7.3 FINDINGS

Strawsonian theories avoid character-related concerns with the help of excuses and exemptions (type-2 pleas or special considerations). Strawsonian theories do not go into detailed expositions of the character traits and their developmental stages, but classify the type of his or her agency, meaning that the quality of agency as a child exempts him or her from responsibility assessments. This derives from the broader scope of morality, which in Strawsonian terms is constituted of the relations between individuals with similar capacities for reason and appraisal.

Wallace's theory in *Responsibility and the Moral Sentiments* (Wallace 1994) is a straightforward extension of Strawson's basic template of moral responsibility. This is particularly apparent in his treatment of the excuses and exemptions, which are a direct elaboration of the special conditions or pleas presented in *FR* (Strawson 2008/[1962], 7-10).

Both character and excuses may indicate instances of *blame mitigation*. In other words, in order to take a wide variety of situations into account the theories of responsibility have incorporated "fail-safes" where the quality of the agent tends (for one reason or another) to invalidate an otherwise valid formalization of an agent acting in a blameworthy manner.

Another way of stating this would be that the cases discussed above which lack the state element, and where the initial responsibility assessment is cancelled based on the these qualities of the agent seemed to indicate a noteworthy feature of the responsibility relations, which the theories seeking to describe moral responsibility should actively acknowledge. The morally sufficient status of the agent should, at least at the most basic level, be verified, so that the theories avoid situations in which an agent – incapable by nature – is unjustly held accountable for wrong-doing. What is interpreted here as missing from most theories of responsibility is a clause about the exemptions that should be a check on whether the moral capabilities of the agent are of a sufficient level.

Excuses are a larger group than the group equivalent to the character, as they take the circumstances unrelated to character into account as well.²⁹¹ This was the case in Strawson's first sub-group of type-2 pleas (see chapter 3.1.3). In Aristotle, the circumstances are sometimes relevant through the character, as with educated virtues (chapter 5.1.2). Otherwise, they are discussed in the voluntary action topic, which is not necessarily related to character. Both have to be taken into account of course in the wider context of responsibility.

An illuminating example of the relationships between the theories can be achieved by cross-referencing the action based-excuses with the character-

²⁹¹ Some authors, such as Michael S. Moore (1990), have even viewed excuses as a higher order category than character-based excuses.

based excuses,²⁹² allowing Aristotle's as well as Strawson's and Wallace's examples of blame mitigation to be compared side-by-side. In all three theories, the category of character-based excuses has the most dramatic effect on the responsibility assessment. Table 7, below, shows this:

Aristotle	Strawson	Wallace
Action-based excuses:	Excusing conditions:	Types of excuse:
a) Involuntariness	1. Unintended or coerced acts (agent normal, less blameworthy)	a) Inadvertance, mistake or accident
a1) Coercion		b) Unintended bodily movement
a2) Ignorance		c) Physical constraint
	2a. Abnormal circumstances (reactive attitudes suspended)	d) Coercion, necessity and duress
Character-based excuses: b) the agent's character belongs in one of the following groups: animals, children	2b. Abnormal psychology (reactive attitudes modified or eliminated)	Exempting conditions: e) the agent is a member of one of the following groups: children, the mentally ill, addicts, the hypnotized, psychopaths or the heavily stressed

Table 7. *Aristotle's, Strawson's and Wallace's groups of excuses compared*

Common to each theory of responsibility are the two broad types of blame mitigation: (1.) the moderate excusing conditions related to slights in action or the circumstances, or (2.) the complete exemption-provoking conditions related to the extraordinary type of agency in question. In the former cases, the initial evaluations of wrong-doing are mitigated to some degree and the

²⁹² The cross-referencing of the action-based excuses with the character-based excuses is similar to the explication in Michael S. Moore's article "Choice, Character and Excuse" (1990). The difference is that Moore's perspective relies more on to the philosophy of law and while organized in a similar fashion, it is not directly related to the comparison done here, but is very interesting nevertheless.

resulting reaction to that event is significantly ameliorated from the uninformed starting position. In the latter, the entire practice of responsibility assessment is halted and cancelled by judgment about the moral level of agency of the agent in question.

Conclusion

Two alternative concepts of responsibility that were discussed were from Peter F. Strawson's and Aristotle's theories of moral responsibility. The elements of responsibility in these theories representing *state* and *response* respectively were special considerations (type-1 and type-2 excuses) and reactive sentiments (including blame) as well as character and praise and, more prominently, blame.²⁹³

It seems that there is a relation in each theory between the intention element and the response element as well as the state element and the response element. The relation between the state and the response elements is that if an agent has an excuse or an exemption in the Strawsonian context, the reactive sentiments related to her action are modified or eliminated. This was made clearer in R. Jay Wallace's extension of Strawson's theory. If a person does not have the character of an ordinary adult human, i.e. does not satisfy the conditions of moral agency the person is not considered morally responsible (subject to praise or blame) in Aristotle's theory.²⁹⁴

Bernard Williams's explication (1993 & 1997) helped to show how factors related to causality (obviously) as well as to intention and state of mind of the agent (as well) can change the resulting blameworthiness. Thus the state of mind is debated in various theories in different terms above as with character in Aristotelian contexts or with the excuses in the Strawsonian ones. If there are interventions to both of these conditions, the initial, reactive evaluation of blameworthiness of the agent can be altered. The central claim here is that the concepts of character and excuses share the connection of being related to the state element of responsibility or that they function as shorthand for the empirical details which might otherwise cause problems for an overly generalized theory of moral responsibility.

²⁹³ The debate is ongoing about whether Aristotle's theory is compatible with Strawson's theory of moral responsibility. As stated, Irwin (1980), Roberts (1989) and Meyer (2011) are discussed in the text. Echenique (2012) and Bobzien (2014) are the latest who hold opposing views on the subject. According to Echenique, Aristotle's blame is not exactly the same as the Strawsonian concept. Bobzien on the other hand emphasizes the collective understanding of the research community that what Aristotle is talking about in *NE* III is recognized as a theory of moral responsibility.

²⁹⁴ The possible exclusions to these are discussed by Strawson's critics, such as Kelly (2013).

8 CONCLUDING STATEMENTS

Finally, after the analysis has been concluded, the results are presented briefly in the form of an argument leading to the actual thesis (8.1) as well as in more detail within the discussion section (8.2).

8.1 THE ARGUMENT OF THIS WORK

Eshleman's structure for theories of moral responsibility is followed here as a guideline.²⁹⁵ Accordingly, the idea or concept of moral responsibility was defined in Strawson's case by showing how the reactive sentiments function and of how moral responsibility should consist of attitudes between the members of the moral community (see *FR*, 23). The criteria for qualifying as a moral agent were defined positively with the participant attitudes and also negatively via the exempted groups, being a member of which invalidated an agent's, responsibility ascription (ibid. 7-11). The conditions under which moral responsibility would be properly applied, i.e., when an agent is responsible for something, was discussed in terms of blameworthiness and the afore-mentioned membership in the moral community in Strawson's theory (and other Strawsonian theories) (see ibid. 23). The objects of responsibility ascriptions in Strawson's case were actions (see ibid. 16).

Aristotle defined the idea or concept of responsibility negatively through the lack of excusing conditions of ignorance and coercion, as well as positively where the agent was a proper candidate for praise and blame (*NE* III.1, 1109b30-35 & 1110b18-20; see also Irwin 1980, 125). The main qualification as a moral agent was being an adult human being with sufficient capacity for making reasoned choices (ibid). Conditions for which moral responsibility would be properly applied related inversely to exclusions of conditions related to being an animal or a child or insane (see *NE* III.2, 1111b8-10). These were discussed in terms of the development of character. Accordingly, the objects of responsibility ascriptions were actions as well as character traits (*NE* III.1, 1109b30-33 & III.5, 1114a6-13; see also Meyer 2011, 122). If we accept that the basic understanding of moral responsibility is in accordance with the Strawsonian reactive sentiments based view of moral responsibility and if Williams's elements are taken to be constitutive of

²⁹⁵ The basic requirements for theories of moral responsibility included (1) the concept, (2) criteria for moral agency, (3) conditions under which responsibility applies to a moral agent and (4) objects of responsibility ascriptions (see Eshleman 2013 & chapter 3.1).

responsibility assessments, meaning that the elements of response (i.e., praise or blame) are recognized as the most apparent manifestation of the responsibility relations, it follows that blame is essential to responsibility. It needs to be noted that the picture of responsibility is incomplete, mainly because of the reluctance of the authors examined here to incorporate the concept of praise into their analysis. Therefore, for the sake of the argument, the praise-related responsibility assessments are not considered further here. (Compare with Coates & Tognazzini 2013, 4-5.)

The overall argument

A) Theories of Responsibility:

1. Agent *s* can be attributed with responsibility and there are ways in which these relations are determined (3.1).
2. Moral responsibility is primarily understood here in terms of Strawsonian sentimentalism: it is constituted from the reactive sentiments, including notions of blame, meaning affective resentment directed at a wrong-doer (3.1.1).
3. While it seems that there can be no single theory of responsibility (following Williams 1993), despite this, accounts of responsibility discussed here contain the Williamsian elements, including most importantly cause, intention, state and response (especially blame) (4.3.3).
4. As a result of 2 and 3, blame is considered to be an essential social component of moral responsibility (3.1, 3.1.2, 5.1.1).

B) Character and Excuses in Theories of Responsibility:

5. If argument A on theories of responsibility is valid, *s* can represent any human being doing act X (3.1).
6. In the situation where X is a defined action, and where the moral status of agent *s* is altered to the extent that blaming the agent is clearly inappropriate, the theory of moral responsibility is in need of a principle of blame mitigation. Otherwise, the truth of the statement is dependent on the capabilities of the agent. (See 3.1.2, 5.1.2.)
7. We can find examples of such principles of blame mitigation in the theories of moral responsibility examined above: Strawson uses special considerations and pleas (3.1.3), which Watson and Wallace interpret as excuses and exemptions (4.1.2, 4.3.4, 6.2), while in the Aristotelian context the cases of children and the insane are discussed in terms of character and

various stages of its development, as well as the excusing conditions of ignorance and coercion. (5.1.2, 6.1).

Conclusion

8. Character and excuses (including exemptions) fulfill the function of blame mitigation in the context of theories of responsibility. This function is needed where responsibility ascription would result in improper outcomes based on the moral status of the agent. These principles, as regulatory components of theories of moral responsibility, have a decisive effect on moral responsibility assessments in general as a result.

8.2 DISCUSSION

Theories of responsibility rarely work with identical concepts or definitions of moral responsibility, a problem corroborated by this research. During the initial research effort, a more particular problem presented itself, of what is the role of character in these theories. By tackling this particularly underdeveloped but increasingly popular area related to responsibility, this second question was selected as the main point of this dissertation. This shift of attention was reflected in the two research questions and the two corresponding parts of the thesis. The first research question (Q1) was solved at least in part by an aporetic conclusion, best worded in Bernard Williams's work on the subject (1993 & 1997; see also chapters 2.2 & 4.3.3). It has been found necessary to discuss the first question in order to address the second properly, as well as to analyze the topic of the latter question.

The second question was raised through the simple observation that the concept of character has received considerably *less* attention in and immediately after the seminal paper on the modern, Strawsonian conception of responsibility, *FR* than it has in connection with the classical texts of Aristotle related to the subject. The theories of responsibility have considered character in different ways throughout the history of Western philosophy, and the ancient Greeks especially saw questions of responsibility for character as an organic part of the interaction between members of society, whereas in modernity the same questions have arisen in a more abstract form.

The relatively recent practice of excluding character from concerns in the theories of responsibility was articulated by comparing the Strawsonian theories with the Aristotelian. For practical reasons, a heuristic approach was selected in that only the theories related directly to the Strawsonian, broadly naturalistic ethics and Aristotle's broadly naturalistic interpretations of responsibility were considered.²⁹⁶

Throughout this comparison it was held that the conceptual tools of the excuses are used in the Strawsonian cases, whereas in Aristotle's case the nature of character is discussed in their place. Strawson and his followers noted character as a cultural factor, or even a qualifier for assessments of moral responsibility altogether, but they did exclude it from the theory of responsibility proper. As a counter-example, McKenna and Russell (2007) noted that the most significant reason for criticism of the default Strawsonian view has been its insufficient treatment of character. In Strawson's case it also is apparent that, while he did eventually help bring about the naturalistic turn, his effort could not ultimately separate the topic of moral responsibility from questions regarding free will and the problem of

²⁹⁶ An example of the latter is Susan S. Meyer's *Aristotle on Moral Responsibility* (2011).

determinism from subsequent philosophical discussions as he might have hoped.

Examination of Aristotle's character virtues provided an alternative viewpoint, one concerned with a bigger picture than just the action evaluated at a given moment. The *Nicomachean Ethics* and *Politics* provide a political theory that has been divided into the two topics by later editorial work. Viewed together, the books present a theory in which questions of responsibility extend from the personal to the political in the form of habituation and education in virtues as well as the relationship between the goals of the actions of the individual and the political ends of the state.

The common denominator between the different theoretical families, *blame*, was used to group the initially unrelated concepts of character and excuse as "forms of blame mitigation"²⁹⁷. A comparison of the two major theories was then conducted (after the more concept-centered systematic presentation), the theories being incorporated into Bernard Williams's suggested framework of elements of responsibility. The comparison showed that the elements of intention and state were those providing interesting distinctions between the theories. While lack of intention seemed to call for re-evaluation of the initial moral responsibility assessments, Williams's state element was often decisive, where the type of agency had a completely eliminative effect on the final response. The latter seemed able to encompass both the Strawsonian type-2 pleas or exemptions and Aristotle's notions of responsibility for character.

As for the question (Q2) about how character should be understood in the context of theories of responsibility, if the argument about the comparability of the various theories of moral responsibility is accepted, it follows that if agent *s* can be any kind of human being doing *X*, situations can be conceived wherein it would be wholly inappropriate to blame *s* for *X*. For example, if an Alzheimer patient who has repeatedly been denied healthcare leaves the kettle on, causing a fire, it would be inappropriate to hold her entirely blameworthy for the damage done despite the causal responsibility being his/hers. If the moral status of the agent is such that expecting flawless behavior under normal circumstances is unrealistic, it can be deemed that the agent is not blameworthy. In cases such as this, it should be asked how the theories of responsibility take extraordinary agents into account. The Strawsonian theories consider them under the special considerations and thus excuse or exempt them from moral responsibility (see chapters 3.1.2, 4.1.2 and 6.2), whereas in the Aristotelian theories character traits determine the moral status of the agent and thus, for example, the young child can be treated differently as still lacking virtues or excellences that are essential for an adult. (5.1.2). Interpreted this way, the concepts of character and excuse

²⁹⁷ In addition to Strawson's and Wallace's precedents, an important influence on this idea was Schoeman (1987).

(in this case including exemptions) are understood as forms of blame mitigation.

In rare cases, character is held to be fundamental in such a way that its presence eliminates the need for a concept of moral responsibility.²⁹⁸ Otherwise, more moderate outlooks can criticize the use of excuses. It seems that while most contemporary theories leave the character-based discussions on responsibility out, they do include type-2 excuses (i.e., exemptions) as descriptors of extraordinary agents (6.2).

Developmental factors and past events mean that philosophical character is “thicker” than psychological character. New entries into empirical-based moral psychology have arrived at conflicting positions with virtue ethics that are influenced by the Aristotelian and other ancient psychological models. Through direct comparisons, the empirically-minded theorists such as Haidt, Prinz, as well as Harman, Doris and Vranas, (Homiak (2015/[2003]) have dismissed the philosophical accounts of character as unrealistic collections of habits and dispositions, almost as horoscopes.

That said the philosophical understanding of character might be too imprecise as a distinct concept to form part of responsibility assessments, which in turn demand precision (see B. Williams 1997). Imprecise, “deep” aspects—such as those championed by Susan Wolf— and by extension responsibility have led the accounts adopting them into a sort of intellectual paralysis of unclarity on whether the agent can be blamed at all. If the environment is considered to be a significant factor in the action of the individual, it might be impossible, now or ever, to blame him. This of course is the point of contention that the hard determinists are after (see Waller 2012).

For this reason the use of character has been associated with incompatibilist accounts of free will and responsibility. The tendency to consider formative events or the past leading to the act in question seems to accompany the argument about determination. This is not the case always however: Watson displays the effect of the “historical” dimension of responsibility in his Strawsonian account (see Watson 1987). Watson does not however advance incompatibilist arguments, as his objective is the sentimentalist formulation of Strawson’s conception of responsibility. By showing that knowledge of character and the formative events of past life affect our moral evaluations (see *ibid.*), Watson criticizes Strawson’s claim that responsibility relations could be construed entirely on affective or emotional grounds.

Furthermore Schoeman (1987) advances character-based ideas in order to question the ledger-view type, overly mechanical attempts at construing responsibility. Finally Fischer and Ravizza’s 1998 theory focusing on reasons-responsiveness, works as an example of a theory that (in Eshleman’s terms) is merit-based and compatibilist (see Eshleman 2014/[2001]). In terms of

²⁹⁸ See Waller 2011.

character, this means that “taking responsibility is a genuine historical notion and its structure is similar to other important historical recursively defined notions” (Fischer & Ravizza 1998, 238). While this might seem to have features familiar in incompatibilist theories, Fischer and Ravizza state that “it is highly attractive to think that taking responsibility is compatible with causal determinism” (ibid. 239).

As based on comparison made here, Aristotle’s responsibility is “more profound” or “thicker” (in Williams’s 1997 terms) than the currently popular Strawsonian theories are. The difference in the role that character plays in Aristotle’s theory explains part of this. The analysis of theories of responsibility through Williams’s categories can help to explain why there are difficulties: the reactive sentiments based approach emphasizes the intention and response elements of responsibility, but they seem to gloss over the effect of the state element through the use of the excuses (and the effect is exceedingly binary in Wallace’s exemptions especially). In modern theories of responsibility, blame mitigation related to the state element has been problematic to pinpoint and to describe, possibly owing to the separation of psychology as its own discipline. In the Strawsonian approach, the cases affected by it are simply excised with the instrument of the excuse, limiting the approach to those that deal with a common adult moral agent. Difficulties with this are related to the cultural definitions of what qualifies as such a moral agent. An Aristotelian character-based approach would produce a similar outcome, in which the differing situation is deemed out-of-context for evaluation, but at least his theoretical context would allow for in-depth explanation of why the agent in question is treated differently. Nevertheless, the problems related to the nature of character would require significant restoration of the core concepts in order for the virtues or some similar functions to be viable.

The authors discussed here who took a positive view of character included Aristotle, who incorporated virtues as states of character (*hexeis*) as central elements of the Aristotelian soul that could guide action. Character was included in Aristotle’s theory of moral responsibility as it was important where responsibility assessments were needed (5.1.2). Indirectly, Smiley’s historical, pragmatic view allowed room for character-based concerns (4.3.2).

A recent source of criticism, the modern psychological theories hailing from situationist disciplines, criticize the Aristotelian notion of character as being insubstantial: all generalizations about the psyche fail given the particular circumstances that force the agent to act apart from them. The most out-spoken opponent of the conception of character and its use in philosophy was John Doris in his multi-disciplinary work *Lack of Character* (2002). The main driver of his argument was based on observations that character is a vague concept and should not be employed as a basis for ethical systems. Doris did, however, allow for the narrow definition for character as a proxy for the everyday psychological evaluations we make of each other to exist (6.1.2). Perhaps, in this sense, Doris can be read in reverse, since his

claim is that character shouldn't be used and we should focus instead on the empirical details. But when the statements about responsibility are made from a normative perspective, there might be a motive to provide general statements. In these cases, character could be interpreted as a way to take into account the cases in which the generalized statement "A is morally responsible for X" clearly fails because of empirical details about A. In other words, character could be read as shorthand for the empirical details in contexts where these generalizations about moral responsibility are sought. Nevertheless, the Aristotelian concept of character is thicker than the psychologists assume: Aristotelian character contains multiple, layered notions of the common psychological patterns of the agent, but also carries the implications of their origin, be it from habituation, education or self-enforced training (see *NE* 1179b20-31). In terms of philosophy, the situationist critique of the Aristotelian character partially misses its mark.

The authors who were both for and against character in different contexts included Bernard Williams, who wrote that if character is interpreted in terms of an element of responsibility – i.e., the *state* element of responsibility, the concept has a crucial role in determining the full picture of responsibility. However, as Williams's more general criticism delineated, moral responsibility itself should not and could not be "profound". Therefore Aristotle's idea about moral education or development forming part of his responsibility of character would be in conflict with Williams's demand (4.3.3 & 7).

Another author who saw both positive and negative sides of the concept was Jean Roberts, whose article on Aristotle suggested that Aristotle's account of responsibility for character was an ill fit with modern accounts of moral responsibility²⁹⁹. On the other hand, she argued that, in Aristotle's context, character is relevant since the agent and her moral education are not simply determined by nature (5.1.2).

The criticisms of the more recent philosophers would strongly suggest that character-based responsibility is inconsistent in itself. Yet even Doris, as the most ardent critic of character-based philosophy is of the mind that there is a pragmatic use for the concept of character within theories of moral responsibility (Doris 2002). Russell (1995), who says Hume is mistaken in emphasising character in responsibility analyses, leaves open the sort of historical responsibility based concerns which would be relevant in determining the lack of a fully functional, adult mind in terms of the concept of character.

As Bernard Williams (1997) allowed for a narrow concept of moral responsibility to be included as moral theory, perhaps in a similar way, a

²⁹⁹ Fischer and Ravizza listed a group of authors who were in agreement and disagreement with Roberts. The other author mentioned was Curren (1989). On the other hand, those authors who thought Aristotle and the modern accounts could be reconciled included T. H. Irwin (1980) and S. S. Meyer (1993). (Fischer & Ravizza 1998, 12, n17.)

limited functionality for character could be left in the conceptual space between full agency and responsible action. At least one generalization seems to arise from this exercise: All theories compared here contain a category of excuses, which relates to the moral qualification of the agent for the responsibility assessments. The concept of character in Aristotle's (and Hume's) theory contributes to these categories. The more recent Strawsonian theories lack this descriptor, but deal with the theme nevertheless through type-2 pleas/special considerations or with the Watsonian/Wallacean exemptions.

By taking note of character, theories of responsibility seem to go further than mechanical responsibility ascription "moment-to-moment" or in terms of Fischer & Ravizza (1998) the "current time-slice" basis. This implies not only that the psychology of the agent matters, but leaves open the potential for social circumstances to have an effect on responsibility assessments historically. As many writers noted, character is – just like responsibility – a social concept. There is no point in employing it on a single individual living in a vacuum.

Following Strawson's example, many have adopted his description of special considerations or pleas as the basis for an alternative way of discussing blame and blame mitigation related reactions (see 3.1.3). Watson and Wallace continued to refine the categories into the clearer definitions of excuses and exemptions (6.2).³⁰⁰ Erin Kelly (2013) was referenced as a philosopher who questions the use excuses in moral theories.

In summary, what was discussed here first was the question of what moral responsibility means in today's philosophy (Q1). After explication of two major theoretical areas of Strawsonian and Aristotelian accounts of responsibility (noting the difficulty of defining them as such), an understanding of the general subject of moral responsibility was established. The position of Bernard Williams and Susan S. Meyer was defended, in that there can be no single, unified theory of moral responsibility. Rather there are discrete elements of responsibility,³⁰¹ and the description of the relation between two of these forms the central thesis of this work.

Williams's elements enabled comparisons between the two very different accounts. The analysis conducted in the seventh chapter displayed a relation between the state- and response elements of both theories. The state element (in the form of Strawsonian exemptions or of similar conclusions derived from Aristotle's character) can in all observed cases have an eliminative or a mitigating effect on the response element described in that theory. Examples included Strawson's special conditions with a "modifying or mollifying" effect on the reactive sentiments, whereas in the Aristotelian picture the

³⁰⁰ Or *mitigating circumstances* and *excuses* if Holly Smith's (2001) definition would have been followed instead (6.2).

³⁰¹ Described by Bernard Williams as including: cause, intention, state and response. For more detail on Williams's argument see chapter 4.3.3.

virtues/vices and the information about the state of character that the act displays to the moral community could similarly lessen the blame (as well as praise) resulting from knowledgeable and uncoerced acts committed by an individual.

Thus, the chapters that have been discussed in connection with each other have focused on the state element (Strawson's special conditions in 3.1.2 and Aristotle's character-based concerns in 5.1.2) and the response element (blame in both, Strawson's reactive sentiments, Aristotle's feedback from the moral community). The claim here, i.e., the thesis, is that in all cases discussed here, the concerns related to the state element of responsibility have had an eliminative or mitigating effect on the response element in the theory in question.³⁰² In other words, the state of the agent (state of mind, character) and the concerns related to it can mitigate or eliminate blame in cases where moral responsibility is being assessed.

As for the answer to the second research question (Q2) "How should character be understood in the context of the philosophy of responsibility?", an answer can be suggested. First, the premise that the theories of responsibility are connected by the notion of blame as the default response to wrong-doing had to be accepted. Given this, an answer could be derived from the application of Bernard Williams' elements of responsibility: character has a definite function within the theories of responsibility and is relevant in cases of extraordinary agency. Naturally occurring instances of blaming can be questioned based on the character in cases of responsibility assessments when the agent is either insufficiently developed in terms of moral capability, or is otherwise prevented from being evaluated as a fully accountable human being.

Character is, according to the commentators, a given pre-requisite (Meyer 2011) or a superficial element in terms of the theories of responsibility for action (Irwin 1980, Roberts 1989), but as a way of bringing an evaluation of the state of the agent into the equation it seems justified. For the idea of moral responsibility, if made clear in all that it implies, the state of the agency must be tested, or at least declared intact. Otherwise if the responsibility assessment is done according to any pre-conceived formula, an agent who is extraordinary in the terms of the state element, compromises the act of assigning moral responsibility. Thus, such checks should be incorporated into the theories of moral responsibility.

What all this implies is that in ascertaining the validity of the responsibility assessments made in each theory here, the agency, or more precisely, the requirements for that agency are evaluated together with the actual responsibility assessment. Discussions of the examples of children and the psychopath have had the following in common: 1) An agent and his or her capability to be held morally responsible is being evaluated. 2) A quality of

³⁰² A similar relation is most likely true between the intention element and the response element, but as the state element is crucial here, this other generalization is left for later attempts.

said agent has a decisive effect on the responsibility assessment related to possible or actual action being committed by said agent. 3) The response in the form of blame is eliminated or at least requires re-evaluation as a result of the quality referred to in 2. In this way, the common area of moral responsibility with extraordinary agents is being discussed in the same format, although with the different conceptual tools related either to the character (and virtues) in the Aristotelian context, or to the special conditions / excuses in the Strawsonian one. Schoeman's question "How do the theories of character and excuses inform one another?" (Schoeman 1987, 8-9) could be answered by comparing how the two theories discuss the cases in which the agent is extraordinary and cannot be held responsible as would a fully responsible, adult moral agent.

One obvious point of difference is the inclusion of moral development that is inherent in the Aristotelian system through the virtues. Virtues, as we saw, were originally habituated or otherwise picked up during upbringing and education. These same virtues in Aristotle's theory have a crucial effect at the time of action, where the important acts are deliberated on, chosen and committed. The way that virtues are conceptualized today could be different, be it in the form of *schemas* or *scripts*, etc. The point of difference nevertheless is that a similar faculty does not exist in the naturalized Strawsonian context. Although character is mentioned in *FR*, it is not developed in relation to the theory of responsibility that follows from the text.

It is argued here that character represents an elaborate or even the most elaborate attempt so far to describe the means of taking historical events into account in assessing moral responsibility. Therefore, even without attempting to re-instate a specific system of character-based ethics, the concept warrants further consideration, especially within the compatibilist camp of moral responsibility theories. Gary Watson's ideas (2004/[1996]) are followed here in that in addition to straightforward accountability for causal responsibility, moral responsibility is considered to contain an element of "aretaic" appraisal within the act of holding one another morally responsible, in cases wherein we enact our ideas about what sort of behaviour we look for in each other, or how we expect to behave in each set of circumstances. As character-inclusive theories have attempted to incorporate this appraisal into the theories of moral responsibility, considerations of character have a property that affects our ascriptions of responsibility, at least in their context.³⁰³ The actual exemptions describe the flip-side of this mechanism: whenever it is deemed within our societies that the agent performing an act otherwise condemnable proves exempt, this same appraisal sets this actor apart from the group of people appropriately blamed for it. In this way, the conceptual wholes of character and excuses are related.

³⁰³ I remain grateful to my pre-examiner Jussi Suikkanen. His questions have helped to clarify these points.

Character additionally needs to be noted in terms of agency. To have a theory of responsibility that would include the inherent limits to the criteria full-fledged moral agency calls for is a requirement, and character should function in this way in the context of theories of responsibility. The worst-case alternative would be that without a clear idea about the agent, assigning responsibility would work only randomly, determined by the qualities of the agent. Without any evaluation of this agent, any theory of responsibility would equate to injustice. If this were to be combined with theories of responsibility centering on action, a modern version of Aristotelian, big-picture evaluation of responsibility could be achieved.

At the least, a test for the “membership of the moral community” should be included in the responsibility assessments in the form of a rudimentary check. If the agent is somehow extraordinary, in terms of the state element (a child or a person who is insane, a psychopath, etc.) it should be immediately clear that further assessment would be impossible. A more detailed, ad hoc evaluations should accompany each responsibility assessment, or the situation would resemble the paradoxical state in which one strives for definitive judgment on one’s responsibility in relation to action and loses the “depth” or profundity of the concept in the process. Otherwise the core idea of responsibility is diluted ad infinitum by having to take every factor of the environment into account. These extremes of overly narrow judgment that is not really moral responsibility as such, and the overly broad, over-determined state of perfect explanation overthrowing any attempt are possibly the only options that we have, but it is hoped that by removing misconceptions about moral responsibility, potentially systematically harmful “solutions” to the problem of moral responsibility can be avoided.

The special conditions or the excuses and exemptions in the Strawsonian context amount to little more than a list of the anomalous agencies, those to which the rules do not apply as they would in perfectly normal cases. It seems that if historical responsibility is to be considered at all, the lack of markers of learning or memory that the virtues represent put the Strawsonian theory in the awkward position of needing to evaluate and qualify the level of moral agency before being able to evaluate the practices and acts being done. This might be overstating the issues of applicability in the Strawsonian approach in general, but the contrast is there.

On the other hand, reversing the question, the theoretization of the excuses is being currently advanced in the Strawsonian context and could as such invigorate the topic of responsibility for Aristotelian virtue ethics as well. A good example of this is Kelly’s paper (2013). The distinctions between the types of agency, and the different levels of blame mitigation between the various types of special consideration could bring much needed detail to the Aristotelian theory, which is only broadly concerned with the voluntariness of the action, including ignorance and coercion as the few excuses discussed in relation to Aristotle’s work. How, for example, would a responsibility assessment sound that could accommodate reasons for exemptions for

specific afflictions or mental conditions that are not discussed in Aristotle, as well as allowing for considerations of moral development akin to the tools that the virtues represent? Unfortunately such a project would be too large to be pursued within the limitations of this dissertation.

Given all this, a cautionary note on generalizing about the use of the concept of character should be sounded here, very much in spirit of Doris (2002). Reincorporating character into modern theories of responsibility would not be a simple task and the reinstatement process would have to happen prudently in that concerns brought by the new scientific advances in data on the mind should not be contradicted. There is significant overlap here with psychological sciences that by many measures are much better able to assemble the facts and truths about the behavior of human beings than the methods of moral philosophy are. Any current ethical theory relying on the concept of character would do well to study these more empirically inclined sciences with care.

It is also worth noting here that the retributive elements of (criminal) responsibility were not in question here. The character-based and excuse-based theories might arrive at precisely the same conclusion. A criminal who is afflicted by the condition of psychopathy (verifiable by some reliable test), can be found guilty and be legally responsible in the same way under both approaches. The only difference is the increased ability of the character-based theory to open up the black box of agency and in this instance inquire about the causes and the circumstances of the wrong-doing. There are philosophers who want their views of responsibility to carry a lawyer-like procession and efficiency. For balance it is hoped that the depth of humanity and history will have some say. Perhaps this is part of “the profound” that Williams was referring to. In that case, it is hoped that clarity could be brought to the matter instead of further complication.³⁰⁴

³⁰⁴ Some of the points of criticism that could be raised based on what has been discussed within this dissertation are anticipated. It is naturally difficult to guess the most pressing concerns of a work during writing, but at least the following difficulties are noted here: First of all, the most obvious criticism that the completed work can face could be due to a crucial *omitted source*. Recognized omissions are acknowledged in footnote 12 in the beginning of chapter 2. A more serious flaw would be a *flaw in argumentation*, and as the argument presented in the previous section relies on certain interpretation of two notoriously imprecise and ambiguous terms – responsibility and character – it needs to be stated with care, that the validity of said argument can be questioned, if a radically different interpretation on one of the two key concepts is made. The specific subject areas of Strawsonian and Aristotelian philosophy, as well as the individual concepts of responsibility, character and excuses are by themselves *too broad*, granted, but as this study has been committed in form of a heuristic, the resulting generalization is within the scope of a single thesis or a claim such as: in the studied theories the identical subject area of blame mitigation is referred to with two different concepts: that of character in Aristotle’s case and that of excuses in the Strawsonian discipline. Furthermore Aristotle and Strawson are seldom discussed head-to-head, with the notable exceptions of Meyer (2011) and Echeñique (2013). Therefore the comparison works as grounds for a unique

perspective, which speaks for the validity of this work. Another possible way of disagreeing with the premises here is to have *different interpretations* of distinct concepts that have been under examination above. Whether the selection done here has been wise or sufficient is up to the reader. Pointing toward an example, the debates about the minutiae of the concept of responsibility and its elements (e.g. blame, see chapter 4.2.) are currently going on, and any input to these debates is welcome. According to other authors, than those that have been examined here, the crucial *combinations of the relevant concepts* and the relations between them could have been significantly different. Treating excuses for example as higher order to character, as Moore (1990) does, could make it difficult to do a comparison like that which has been done above. Here moral responsibility and theories concerning it is treated in higher hierarchical order than the concepts of character and excuses, which are considered here to be on equal footing. Finally, regarding *skepticism* of moral responsibility, the most direct opposition to the entire idea has been advanced in the form of hard determinism. Pereboom (see 2001) might be the most recognized figure related to the subject, as is Waller's fairly recent work (2011). Because the hard determinist position demands that the practice of blaming should be ceased, at least in the sense that the sentiment would not lead to social sanctioning and because blame is integral to the practices related with moral responsibility, which following the broadly naturalist outset that was chosen as the common ground for this multi-disciplinary effort, allows for it and presents viable descriptions of the actual, social interaction between the human beings, the former position is deemed incompatible with the results arrived at here. For this reason the principles or the position of hard determinism cannot be accepted from the perspective of the argument. However, it is not wished that dialogue should end. On the contrary any comments from the different viewpoints are welcome. A more moderate form of skepticism is, however, called for: Following Williams (1993), even as there can be no overarching theory, any account of moral responsibility cannot be complete without all of its elements, including those related to character.

9 CONCLUSION

The concept of responsibility and its discussion in the Anglophone philosophical world was anchored here on Peter Strawson's article "Freedom and Resentment" (2008/[1962]). The discussion on moral responsibility, as observed in that article, was shifted away from the fact-value debates favored by the logical positivists toward a more sentimentalist view emphasizing naturalistic moral psychology instead of linguistics.

The work of the preeminent contemporary Strawsonian theorist R. Jay Wallace (1994) was examined in order to formulate a heuristic idea of how Strawson's views have developed further in our own era. The conclusion of the initial examination was that even if Strawson's work had considerably advanced how interesting a concept responsibility is, and his progress toward the sentimentalist outlook made the discussion relevant again, on the downside the free will vs. determinism debate still hindered the debate considerably, especially concerning the applicability of the theories of responsibility.

So far, it has been exceedingly difficult to pinpoint moral responsibility as a precise concept in the history of philosophy. While the naturalism-based viewpoints have made the topic accessible by letting the writers concentrate on the reactive sentiments and common concepts such as praise and blame as the distinguishing qualities of responsibility relations constituting the idea of moral responsibility, critical examinations of the subject cannot be ignored altogether. A frequent criticism of Strawson and his followers was the insufficiency of their accounts of events in history that had an effect on the responsibility assessments of a particular moment.

Criticisms of moral responsibility by Bernard Williams (1993 & 1997) were addressed, the most poignant message from him being that there could never be a unified theory of responsibility, because the concept contains four distinct elements: cause, intention, state and response. These elements appear in different configurations whenever moral responsibility is discussed. Moral responsibility includes, in addition to mere retrospective responsibility for action, other notions worth considering: being responsible and what it means was a problem that occupied writings such as those of Herbert Fingarette (1966) and Gary Watson (1987). The latter touched on the theme of historical responsibility by looking at acts which were not done precisely in connection with responsibility assessments, but earlier, in formative ways (see Watson 1987 & chapter 3.4.4). Marion Smiley (1992) presented even more of an objective assessment of responsibility by making a historical comparison of the modern (Strawsonian) and classical (Aristotelian) accounts of responsibility, which were the two main theories compared in this study.

Aristotle's conception of responsibility, which was considered in chapter 5, was given three interpretations. The first of these, responsibility for action, was mostly informed by Terence H. Irwin (1980), who presented Aristotle's theory in the form of the "simple theory" ("A is responsible (a candidate for praise or blame) if A does X voluntarily").³⁰⁵ Historical responsibility in his theory led Irwin to formulate a "complex theory", which included consideration of capacity to make *prohairesis*-type informed decisions. Political responsibility was briefly looked at, but was ultimately not included.³⁰⁶ Among others, Jean Roberts (1989) did not accept the unifying theory and, along with Susan Meyer (2011), she considered Aristotle's responsibility for character as a separate, albeit questionable theme. Meyer claimed that the character-based writings found in book III.5 of the *Nicomachean Ethics* were not part of Aristotle's moral responsibility, being part of another argument with more to do with moral education.

The definition of responsibility as preparedness for blame and sanction has been refuted on many occasions since. In this connection, the key problem that this work has been insufficient in scope to solve, is that by following Williams's 1993 theses, philosophical accounts of responsibility are incapable of tackling the "deep" intentions associated with the concept. This means that talk about responsibility must be kept sufficiently "narrow" in terms of definitions. The resulting narrow yet functional theoretizations (such as Wallace's) have not met with universal acceptance³⁰⁷, nor have the theories that seek the deeper elements (by Wolf, Smiley, etc.) succeeded in incorporating all that is necessary in a manner that could be agreed upon. Taking into account Bernard Williams's claim (1997) that moral responsibility cannot be "profound", as well as Watson's 1987 example of historical effect on the reactive sentiments; as well as the unease of the aforementioned Aristotle scholars on the topic of responsibility for character (5.2), it was deemed that character in theories of responsibility was worth a closer, systematic look.

The second part of the dissertation consisted of the systematic examination of the various definitions and uses of the concepts of character and excuses as well as their analysis. The first definition was discussed in the

³⁰⁵ Whether or not Irwin's addition to Aristotle's theory of responsibility in the form of his "complex theory" was correct or not, it is considered here that in responsibility for action his "simple theory" was adequate, and has frequently been referred to since. (Irwin 1980, 125, chapter 3.1.1.)

³⁰⁶ Political responsibility as understood by Delba Winthrop (1975) and others was taken into account as a connected relevant theme. However, as the core functioning of interpersonal responsibility, as understood by proponents of the naturalistic interpretation (which was chosen as the primary discipline, in order for the multi-disciplinary study to be possible), did not include political responsibility, at least directly, it was decided not to examine the third category of Aristotle's responsibility further.

³⁰⁷ Compare, for example, with Paul Russell's critique of Wallace presented in his article "Responsibility, Naturalism, and the Morality System" (2013).

case of Strawson and his followers in only a limited way, noting how their concept was usually all but dismissed and the topic was relegated with the help of the excuses.

Character, as it appeared in Aristotle's theory was fleshed out further with the help of texts by authors such as Broadie (1991) and Burnyeat (1999), among others. The concept was included by default in the considerations of responsibility for action, and had an effect on whether the act could be considered blameworthy or not depending on the state of the character of the agent. This was found to be especially interesting where the agent was considered extraordinary, as in not a fully developed adult human being. The basic formulation of "s is responsible for X" was found alterable in terms of the response element, blame, by substitution of different types of agents. If s was an infant, for example, or a mentally incapacitated adult, the outcome of the responsibility assessment would be different than where s was an ordinary adult human being. Character, while advocated by some authors like McIntyre, has received more dismissal and criticism, as from John Doris, who viewed the concept as both outdated psychology and as a misleading umbrella-term. Doris's psychologism was discussed critically as well.

In the corresponding modern, Strawsonian theories (at least as interpreted by Watson and Wallace) the similar arguments concerning the relation between state of the agent and response to the act were not discussed in terms of character. Instead, these propositions in Strawson's article were discussed in terms of pleas or special considerations and later by his followers in terms of excuses (called exemptions in the most severe instances).

The contrast between character and excuses was used to highlight how the theories of responsibility deal with with extraordinary agents. Strawson equated the response to acts by children in the same dismissal categories as persons with serious personality-altering disabilities, etc. Similarly, Wallace posited unintuitive concepts as "blameworthiness inhibitors" in order to explain why the moral responsibility of some people was treated differently.

This contrast was explored by looking at theories of responsibility. The modern, default way of discussing blame mitigation seemed to be centered on excuses, while character has been avoided as a concept, at least by the philosophers in the compatibilist camp. As for the argument of the dissertation, it seemed clear at this point that the concerns of character were an element of theories of responsibility from Kant onward right up to the modern philosophers, and one reason that might explain why contemporary philosophers are so reluctant to discuss character is that psychology has become its own science and these questions could be regarded as belonging in that empirical setting. Nevertheless, the evidence seemed to support a need for re-evaluation of character in philosophical theories of responsibility. The naturalistic approach has deliberately blurred the boundaries between philosophy and psychology since Quine. After this, different criticisms were addressed.

The final question was posed in terms of whether character as a concept has a role in modern theories of responsibility, and the question was analyzed by comparing the Williamsian elements of responsibility. The question of whether Aristotle's and Strawson's theories of responsibility are compatible was answered by means of comparing the categories of excused and exempted agents. The state element was taken to signify settings similar to them and those discussed in terms of the Aristotelian topic of responsibility for character in *Nicomachean Ethics* III.5. While some writers viewed character in theories of responsibility as borderline "occult qualities", following Hume, even the most ardent critics (especially Doris 2002) recognized the pragmatic everyday use of the concept as a simplifying tool of evaluation. This permits a role for character in modern theories of responsibility as well, and this use is twofold:

1. Character as a philosophical concept allows condensation of the effect of historical responsibility of the agent into easily understandable terms in cases requiring responsibility assessment. By way of prospective assessment, it allows the simple device of anticipating the probable course of action taken by the agent in advance, and with this a more realistic account of interaction between the agents, such as is relevant in the current debate on "standing to blame" (see chapter 4.3.4).

2. In addition to this, character can in any social circumstance function as a simple summary of the mental state of the agent at the time of the act. If incorporated into theories of responsibility, this would avoid the need for convoluted conditions for how the theory should take various types of extraordinary agents into account (chapters 5.1.2 & 6.1).

Excuses, while a broader category, can incorporate the same functions:

3. Excuses can mitigate blame moderately or completely, at which point they are called exemptions (chapter 6.2).

4. Excuses can incorporate the points 1 and 2 as character-based excuses (see table 6 in chapter 7.3). However, this is seldom acknowledged within the existing theories and even there it is done relatively superficially in contrast to the myriad factors needed in character formation in the theories favouring the latter concept.³⁰⁸

Any theory leaving 1 and 2 out of its scope can arrive at a situation in which the grounds for excusing are insufficiently defined. While Strawson did mean his theory to take into account the social context or "the web of human

³⁰⁸ For ways that Aristotle's character has an effect on excuses (see Broadie 1991, 159-174). Concerning Hume, the Hume-Brandt thesis, discussed by Holly M. Smith also combines the notions of character with the excuses (see 2001, 507).

attitudes and feelings” and despite the departure from the norm being significant from his contemporary perspective, it seems that the difference between his and Aristotle’s theory is a step away from the pragmatic reality. Aristotle meant “his account to be relevant to ordinary judgments about responsibility.” (Irwin 1980, 120.) His theory, as interpreted by the contemporary scholars is, more so than Strawson’s, a theory of moral responsibility *meant to be applied*. Strawson did not sufficiently follow up the initial masterstroke of “Freedom and Resentment”, but the work of his followers, as evidenced by contributions up to the recent series of *Oxford Studies in Agency and Responsibility* (Shoemaker (ed.) 2013, 2014 (with Tognazzini, eds.) & 2015) show that the potential has always been there for a complete and intricate theory, perhaps so far best realized by Wallace in his *Responsibility and the Moral Sentiments* (1994).

Looking at the notion of character in the context of philosophy of responsibility at least showed that there is room for expansion in the narrower attempts, i.e., those relying on excuses, in defining the content and limits of moral responsibility. Even without demands to bring back Aristotle’s complex, yet outdated ideas about character (as described by Irwin 1980, Roberts 1989 and Meyer 2011), the considerations related to that theory (see Fingarette 1966, Watson 1987 & Fischer and Ravizza 1998) describe facets of responsibility, and indeed a form of moral agency, that proponents of the currently popular Strawsonian theories (Wallace (1994) at the forefront) could pay more attention to. As an example, the notion of historical responsibility (as discussed in Watson (1987) and Fischer & Ravizza (1998)) seems a good candidate for this, as are the recent initiatives on the topic of “standing to blame” (see Bell 2013). As a result, the way is open to the constant evolution of contending accounts of blame and other related topics. Whatever direction the philosophy of responsibility does take in the near future, it is certain that the debate is going to be heated, lively, and captivating.

BIBLIOGRAPHY

Classics

- Aristotle. 1984. *The Collected Works of Aristotle*. (incl. *NE* & *Pol.*) Edited by Jonathan Barnes. Princeton University Press, Princeton. (*Nicomachean Ethics*, *Politics* and *Rhetoric* quotes in text are from this edition.)
- , 1990. *Aristoteles VI: Metafysiikka*. Finnish translation by T. Jatakari, K. Näätsaari & P. Pohjanlehto. Gaudeamus. Helsinki.
- , 1991. *Aristoteles VIII: Poliitiikka*. Finnish translation by A. M. Anttila. Gaudeamus. Helsinki.
- , 1998. *Politics*, translated by Ernest Barker. Revised and notes by R. F. Stalley. (2nd ed.) Cambridge: Cambridge University Press.
- , 2000. *Nicomachean Ethics*, translated by Roger Crisp. Cambridge: Cambridge University Press.
- , 2005. *Aristoteles VII: Nikomakhoksen etiikka*. Finnish translation by Simo Knuuttila. 2005 (2nd printing) Gaudeamus. Helsinki.
- , 2011. *The Eudemian Ethics*, translated by Anthony Kenny. Great Britain: Oxford University Press. (Quotes in text are from this edition.)
- Hume, David. 2000/[1739–40]. *A Treatise of Human Nature*, (THN) David Fate Norton and Mary J. Norton (eds.), (Oxford: Oxford University Press).
- , 2000/[1748]. *An Enquiry Concerning Human Understanding*, (ECHU) Tom L. Beauchamp (ed.), (Oxford: Oxford University Press.). Finnish translation: *Tutkimus Inhimillisestä Ymmärryksestä* by Eino Kaila, 1938. Porvoo: WSOY.
- Plato, 1961. *The Republic*. From *Plato: The Collected Dialogues*. Edited by Edith Hamilton and Huntington Cairns. Bollingen Foundation. Princeton, New Jersey: Princeton University Press. Finnish translation: *Valtio* by Marja Itkonen-Kaila, 2007. Otava.
- Strawson, Peter F. 2008/[1962]. “Freedom and Resentment”. (FR) In Strawson, P. F. 2008. *Freedom and Resentment and Other Essays*. Routledge, Abingdon. Originally in *Proceedings of the British Academy* 48, 1-25.
- , 2008. *Freedom and Resentment and Other Essays*. Routledge, Abingdon.

Commentators

- Alfano, M. 2013a. "Identifying and Defending the Hard Core of Virtue Ethics" in *Journal of Philosophical Research* 38:233-260.
- Anscombe, G. E. M. 1958. "Modern Moral Philosophy". *Philosophy*, 33: 1–19.
- , 1977. "Thought and Action in Aristotle". In the book J. Barnes, M. Schofield & R. Sorabji (eds.) *Articles on Aristotle 2. Ethics and Politics*, Surrey: Duckworth.
- Applbaum, A. I. 1994. [Review of Smiley's] "Moral Responsibility and the Boundaries of Community" (1992). *Political Theory*, Vol. 22, No. 21 (Feb. 1994), pp. 186-190.
- Austin, J. L. 1956-7. "A Plea for Excuses: The Presidential Address" *Proceedings of the Aristotelian Society*, New Series, Vol. 57 (1956 – 1957), pp. 1-30.
- Badhwar, N. 2009 "The Milgram Experiments, Learned Helplessness, and Character Traits" *Journal of Ethics* 13 (2-3):257 - 289. (<http://link.springer.com/article/10.1007%2Fs10892-009-9052-4>) Read 5.10.2015
- Bayles, Michael D. 1976. "Hume on Blame and Excuse" *Hume Studies* Vol. 2, no. 1 (April 1976), 17-35.
- Bazargan, S. 2014. "Moral Coercion" *Philosophers' Imprint* vol. 14, No. 11 (May 2014), .
- Becker, L. C. & Becker, C. B. (eds.) 2001. *Encyclopedia of Ethics*. 2nd edition. London: Routledge.
- Bobzien, Susanne 2014. "Choice and Moral Responsibility in Nicomachean Ethics iii 1-5". In *The Cambridge Companion to Aristotle's Nicomachean Ethics*. R. Polansky (ed.) Retrieved from academia.edu (20.4.2015).
- Broadie, Sarah 1991. *Ethics with Aristotle*, New York: Oxford University Press.
- Brink, D. O. & Nelkin D.K. 2013. "Fairness and Architecture of Responsibility" in Shoemaker (ed.) 2013.
- Burnyeat, M. F. "Aristotle on Learning to Be Good" in Sherman, N. 1999. *Aristotle's Ethics. Critical Essays*. (Also in A. Oksenberg Rorty (ed.). 1980.)
- Butler, Joseph 1849. "On Resentment" [Sermon 8], in *The Works of Joseph Butler*, Vol. 2. *The Works of Joseph Butler*, 2 vols, Samuel Halifax (ed.), (Oxford: Oxford University Press, 1849).
- Campbell, C. A. 1951. "Is 'Freewill' a Pseudo-Problem?", *Mind*, 60, 446–65. Reprinted in Bernard Berofsky, (ed.), *Free Will and Determinism*.
- Chappell, T. D. J. 1996. [Review of S.S. Meyer's] "Aristotle on Moral Responsibility" *Mind*, Jan. 1996.
- Clarke, R., McKenna, M. & Smith, A. 2015. *The Nature of Moral Responsibility - New Essays*. Oxford: Oxford University Press.
- Coates, Justin & Tognazzini, Neal (eds). 2013. *Blame. Its Nature and Norms*. Oxford: Oxford University Press.
- Cohen, G. A., 2006. "Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists?" *Royal Institute of Philosophy Supplement*, 58: 113-136.
- Curren, R. R. 1989. "The Contribution of *Nicomachean Ethics* iii 5 to Aristotle's Theory of Responsibility" in *History of Philosophy Quarterly*. Vol. 6 no. 3 (July 1989).

- Darwall, S. 2006. *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Duff, A. 1998. "Responsibility" in *Routledge Encyclopedia of Philosophy*, E. Craig (ed.) Routledge.
- Doris, John. 2002. *Lack of Character. Personality and Moral Behaviour*. Cambridge, Massachusetts: The University of Cambridge Press.
- , 1998. "Persons, Situations, and Virtue Ethics." *Nous* 32: 504–540. (<http://onlinelibrary.wiley.com/doi/10.1111/0029-4624.00136/abstract;jsessionid=F76EE0E1FFC96E87706E5ED67E99CB51.f01t03>) Read 5.10.2015.
- Echeñique, J. 2012. *Aristotle's Ethics and Moral Responsibility*. Cambridge: Cambridge University Press.
- Eshleman, Andrew S. 2014b. "Worthy of Praise" in Shoemaker, D. & Tognazzini N. (eds.), 2014. *Oxford Studies in Agency and Responsibility (Volume 2) 'Freedom and Resentment' at 50*.
- Feinberg, Joel. 1970. *Doing and Deserving*. Princeton: Princeton University Press.
- Fingarette, Herbert. 1966. "Responsibility". In the publication: *Mind*, New Series, Vol. 75, No. 297 (Jan., 1966), s.58–74.
- Fischer, John M. & Ravizza, Mark 1998. *Responsibility and Control. A Theory of moral Responsibility*. Cambridge: Cambridge University Press.
- Fischer, John M. & Ravizza, Mark (eds.), 1993. *Perspectives on Moral Responsibility*. Ithaca: Cornell University.
- Fischer, John M. 1987. "Responsiveness and moral responsibility" in Schoeman (ed.) 1987.
- , 1982. "Responsibility and Control", *Journal of Philosophy*, 89 (1982), 24–40.
- Frankena, William 1973. *Ethics*. 2nd edition. New Jersey: Prentice Hall.
- Godman, M. & Jefferson, A. 2014. "On Blaming and Punishing Psychopaths". *Criminal Law and Psychology*. Published online 4.10.2014. Springer.
- Gracia, Reichberg & Schumacher (eds.) 2003. *The Classics of Western Philosophy A Reader's Guide*. Blackwell Publishing, Padstow, Cornwall.
- Harman G. 2009. "Three Trends in Moral and Political Philosophy." *The Journal of Value Inquiry* 37: 415–425. (<http://link.springer.com/article/10.1023%2FB%3AINQU.0000013350.42507.c6>) Read 5.10.2015.
- , 2003, "No Character or Personality." *Business Ethics Quarterly* 13: 87–94. (<http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=9460471&fileId=s1052150X00002116>) Read 5.10.2015.
- , 2000, "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society* 100: 223–226. (<http://onlinelibrary.wiley.com/doi/10.1111/j.0066-7372.2003.00013.x/abstract>) Read 5.10.2015.
- , 1999. "Moral Philosophy Meets Social Psychology." *Proceedings of the Aristotelian Society*, New Series 119: 316–331
- Hart, H. L. A. 1968. *Punishment and Responsibility*. New York: Oxford University Press.
- , 1968. "Legal Responsibility and Excuses" in Hart H.L.A. 1968 *Punishment and Responsibility*.
- Hieronimi, P. 2008. "Sher's Defense of Blame" *Philosophical Studies*, 137(1): 19–30.

- , 2004 “The Force and Fairness of Blame.” *Philosophical Perspectives* 18:115–48.
- Hurka T. 2006. “Virtuous Act, Virtuous Dispositions.” *Analysis* 66: 75. (<http://analysis.oxfordjournals.org/content/66/1/69>) Read 5.10.2015
- Irwin, T. H. 1980. “Reason and Responsibility in Aristotle”, in the book Amélie Oksenberg Rorty (ed.) *Essays on Aristotle’s Ethics*, Berkeley: University of California Press.
- , 2001. “Aristotle” in Becker, L. C. & Becker C. B. 2001. *Encyclopedia of Ethics*. London: Routledge.
- , 2003. “Aristotle, *Nicomachean Ethics*. (367 – 323 bc) A Sort of Political Science” in *The Classics of Western Philosophy A Reader’s Guide*, Gracia, Reichberg & Schumacher (eds.), Blackwell Publishing, Padstow, Cornwall.
- Kaila, Eero. 2011. ”Teot, luonne ja kansalaisuus, Nykyfilosofian tulkintoja Aristoteleen vastuukäsityksestä”, *Historiallinen aikakauskirja* 2/2011, 171–179.
- , 2014. ”Luonteen hyveet valtion tasolla: Platon ja Aristoteles” in *Hyve*. Hämäläinen, N., Lemetti, J. & Niiniluoto, I. (eds.) *Filosofisia tutkimuksia Helsingin yliopistosta*. Vol. 45. Helsinki: University of Helsinki.
- Kamtekar, R. 2004. “Situationism and Virtue Ethics on the Content of our Character.” *Ethics* 114: 458–491. (<http://www.jstor.org/stable/10.1086/381696>) Read 5.10.2015
- Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kelly, Erin, I. 2002. “Doing Without Desert”, *Pacific Philosophical Quarterly*, 83, 180–205.
- , 2013. “What Is an Excuse?” in Coates, Justin & Tognazzini, Neal (eds). 2013 *Blame. Its Nature and Norms*.
- Kennett J. 2008. “Reasons, Reverence and Value” in Sinnott-Armstrong (ed.) 2008.
- Kinnaman, T. 2005. “The Role of Character in Hume’s Account of Moral Responsibility”. *The Journal of Value Inquiry*. (2005) 39: 11-25.
- Knuuttila, Simo 1998. “Aristoteles” in Korkman, P. & Yrjönsuuri, M. (ed.) 1998. *Filosofian historian kehityslinjoja*.
- Korkman, P. & Yrjönsuuri, M. (ed.) 1998. *Filosofian historian kehityslinjoja*. Tampere: Gaudeamus.
- Korsgaard, C. 1996. “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations” in Korsgaard, Christine 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Kupperman J. J. 2009. “Virtue in virtue ethics”. *Journal of Ethics* 13 (2-3): 243-255.
- Kutz, C. 2002. “Responsibility” in J. Coleman & S. Schapiro (eds.) *Oxford Handbook of Jurisprudence and Philosophy of Law*
- MacIntyre, A. 1981. *After Virtue: A Study in Moral Theory*. Notre Dame, IN: University of Notre Dame Press.
- Macnamara, C. 2013. “Taking Demands Out of Blame”. in Coates, J. & Tognazzini, N. (eds.). 2013 *Blame. Its Nature and Norms*.
- McGeer, V. 2008a. “Varieties of Moral Agency: Lessons from Autism (and Psychopathy)” in Sinnott-Armstrong (ed.) 2008.
- , 2008b. “The Makings of a Moral Sensibility: Replies to Commentaries” in Sinnott-Armstrong (ed.) 2008.

- , 2013. “Civilizing Blame” in Coates, J. & Tognazzini, N. (eds.). 2013 *Blame. Its Nature and Norms*.
- , 2014. “P. F. Strawson’s Consequentialism” in Shoemaker, D. & Tognazzini N. (eds.), 2014. *Oxford Studies in Agency and Responsibility (Volume 2) ‘Freedom and Resentment’ at 50*.
- McIntyre, Jane L. 1990. “Character: A Humean Account”. *History of Philosophy Quarterly*. Vol. 7, no. 2, April 1990.
- , 2015. ““Repentance wipes off every crime”: Character, the self and moral responsibility in Hume’s philosophy”. Conference presentation delivered at The 42nd International Hume Society Conference at Stockholms universitet (23.7.2015). Referred to with permission by the author.
- McKenna, M. 1996. Review “Responsibility and the Moral Sentiments” by R. Jay Wallace. *The Philosophical Review*, Vol. 105, No. 3 (Jul. 1996), pp. 415-418.
- , 1998. “The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism” *Journal of Ethics* 2, 123-142.
- , 2005. “Where Strawson and Frankfurt Meet”, *Midwest Studies in Philosophy*, 29 (2005), 163–80.
- , 2013. “Reasons-Responsiveness, Agents, and Mechanisms”, in Shoemaker, David, ed., 2013. *Oxford Studies in Agency and Responsibility (Volume 1)*, New York: Oxford University Press.
- McKenna, M. & Russell, P. (eds.) 2007. *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson’s “Freedom and Resentment”*. Aldershot: Ashgate.
- Mele, A. 2009. *Effective intentions: The power of conscious will*. New York: Oxford University Press.
- Merritt M. 2000. “Virtue Ethics and Situationist Personality Psychology.” *Ethical Theory and Moral Practice* 3: 365–383. (<http://link.springer.com/article/10.1023%2FA%3A1009926720584>) Read 5.10.2015
- Meyer, S. S. 2011/[1993]. *Aristotle on Moral Responsibility: Character and Cause*. 2nd edition. Oxford: Blackwell.
- Milgram, S. 1963. “Behavioral study of obedience”. *Journal of Abnormal and Social Psychology* 67, 371-378.
- Miller, C. 2009. “Empathy, Social Psychology, and Global Helping Traits.” *Philosophical Studies* 142: 247–275 (<http://link.springer.com/article/10.1007%2Fs11098-007-9185-x>) Read 5.10.2015
- , 2013 *Moral Character. An Empirical Theory*. Oxford: Oxford University Press.
- Miller, Furr, Knobel & Fleeson (eds.) 2015. *Character. New Directions from Philosophy, Psychology, and Theology*. New York: Oxford University Press.
- Monroe, K. R., Martin, A. & Ghosh, P. 2009. “Politics and an Innate Moral Sense: Scientific Evidence for an Old Theory?” *Political Research Quarterly*, Vol. 62, No. 3 (Sep., 2009), pp. 614-634. (JStor, 19.8.2011).
- Moore, Michael S. 1990. “Choice, Character and Excuse”. *Social Philosophy and Policy* 7 (02).
- Mäkelä, P. 2013. *Collective Responsibility: Against Collectivism*. Helsinki: University of Helsinki
- Nichols, Mary P. 1992. *Citizens and Statesmen*. Savage: Rowman & Littlefield Publishers Inc.

- Nowell-Smith, H. P. 1948. "Free Will and Moral Responsibility", *Mind*, 57 (225), 45– 61.
- , 1960. "Ifs and Cans", *Theoria*, 26 85–101.
- Nussbaum, M. 1986. *The Fragility of Goodness*. Cambridge: Cambridge University Press.
- Oksenberg Rorty, Amélie (ed.) 1980. *Essays on Aristotle's Ethics*, Berkeley: University of California Press.
- Pavlopoulos, M. 2003. "Aristotle's Natural Teleology and Metaphysics of Life" in *Oxford Studies in Ancient*. Vol. XXIV (Summer 2003), D. Sedley (ed.). Oxford University Press.
- Pereboom, D., 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Prinz, Jesse J. 2010. "The Roles of Emotion in Morality" *The Oxford Handbook of Philosophy of Emotion*. Oxford Handbooks Online: January 2010.
- , 2011. *The Emotional Construction of Morals*. (2nd print). Oxford: Oxford University Press.
- Polansky, R. (ed.) 2014. *The Cambridge Companion to Aristotle's Nicomachean Ethics*. R. Polansky (ed.) Cambridge: Cambridge University Press.
- Pritchard, M. S. 1991. *On Becoming Responsible*. Lawrence, Kansas: University Press of Kansas.
- Roberts, Jean 1989. "Aristotle on Responsibility for Action and Character", *Ancient Philosophy* Vol. IX No. 1.
- Russell D. 2009 *Practical Intelligence and the Virtues*. Oxford: Oxford University Press.
(<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199565795.001.0001/acprof-9780199565795>) Read 5.10.2015
- Russell, Paul 1992. "Strawson's Way of Naturalizing Responsibility" *Ethics* 102, 287-302.
- , 1995. *Freedom and Moral Sentiment: Hume's way of Naturalizing Responsibility* New York: Oxford University Press.
- , 2013. "Responsibility, Naturalism, and "The Morality System"" in Shoemaker, David, ed., 2013. *Oxford Studies in Agency and Responsibility (Volume 1)*, New York: Oxford University Press.
- Salkever, Stephen G. 1990. *Finding the Mean*, New Jersey: Princeton University Press.
- Sarkissian, H. 2010 "Minor Tweaks, Major Payoffs." *Philosopher's Imprint* 10.9: 1–15.
- Scanlon, Thomas M. 1986. "The Significance of Choice". *The Tanner Lectures on Human Values VIII*. Salt Lake City: University of Utah Press.
- , 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- , 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap.
- , 2013. "Interpreting Blame" in Coates, J. & Tognazzini, N. (eds.). 2013 *Blame. Its Nature and Norms*.
- Scheffler, Samuel 1992. "Responsibility, Reactive Attitudes, and Liberalism", *Philosophy and Public Affairs*, 21 (1992), 299–323.
- Schlick, Moritz, 1966/[1939]. "When Is a Man Responsible?" reprinted in Berofsky (ed.), *Free Will and Determinism*.

- Schoeman, Ferdinand (ed.) 1987. *Responsibility, Character and the Emotions. New Essays in Moral Psychology*. Cambridge: Cambridge University Press.
- , 1987. "Statistical norms and moral attributions" in Schoeman, Ferdinand (ed.) 1987. *Responsibility, Character and the Emotions. New Essays in Moral Psychology*.
- Sher, G. 2006. *In Praise of Blame* Oxford: Oxford University Press.
- , 2013. "Wrongdoing and Relationships: The Problem of the Stranger" in Coates & Tognazzini (eds.) 2013.
- Sherman, N. 1999. *Aristotle's Ethics. Critical Essays*. Oxford: Rowman & Littlefield Publishers Inc.
- , 1999a. "The Habituation of Character" in Sherman 1999. *Aristotle's Ethics. Critical Essays*.
- , 2001. "excellence" in Becker, L. C. & Becker C. B. 2001. *Encyclopedia of Ethics*. London: Routledge.
- Shoemaker, David, (ed.), 2013. *Oxford Studies in Agency and Responsibility (Volume 1)*, New York: Oxford University Press.
- , 2013. "Introduction". In Shoemaker, David, (ed.), 2013. *Oxford Studies in Agency and Responsibility*.
- , 2015 (ed.). *Oxford Studies in Agency and Responsibility (Volume 3)*, New York: Oxford University Press.
- Shoemaker, D. & Tognazzini N. (eds.), 2014. *Oxford Studies in Agency and Responsibility (Volume 2) 'Freedom and Resentment' at 50*, UK: Oxford University Press.
- Sinnott-Armstrong, W. (ed.) 2008. *Moral Psychology volume 3. The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge MA: The MIT Press.
- Smart, J.J.C. 1961. *An Outline of a System of Utilitarian Ethics*. Melbourne: Melbourne University Press & Toronto: Macmillan.
- , 1963. "Free Will, Praise, and Blame", *Mind*, 70 (1963), 291–306.
- Smilansky, Saul. 2000. *Free Will and Illusion*. Oxford: Oxford University Press.
- , 2001. "Free Will: From Nature to Illusion", *Proceedings of the Aristotelian Society*, 101, 71–95.
- Smiley, M. 1992. *Moral Responsibility and the Boundaries of the Community. Power and Accountability from a Pragmatic Point of View*. The University of Chicago Press, Chicago.
- Smith, Angela M. 2007. "On Being Responsible and Holding Responsible," *The Journal of Ethics*, 11: 465–484.
- , 2013. "Moral Blame and Moral Protest" in in Coates, Justin & Tognazzini, Neal (eds.). 2013 *Blame. Its Nature and Norms*.
- Smith, Holly M. 2001. "excuses" in Becker, L. C. & Becker C. B. 2001. *Encyclopedia of Ethics*. London: Routledge.
- Snowdon, P. 2009. "Peter Frederick Strawson", *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2009/entries/strawson/>>. (20.8.2014).
- Sorabji, Richard 1980. "Aristotle on the Role of Intellect in Virtue", in the book Amélie Oksenberg Rorty (ed.). *Essays on Aristotle's Ethics*, Berkeley: University of California Press.

- Sreenivasan G. 2002 "Errors about errors: Virtue theory and trait attribution". *Mind* 111 (441):47-68.
- Todd, P. 2012. "Manipulation and Moral Standing: An Argument for Incompatibilism" [vol. 12, no. 7, pp. 1-18]
- Wallace, R. J. 1994. *Responsibility and the Moral Sentiments*. Harvard University Press, Cambridge, Massachusetts.
- , 2002. "Scanlon's Contractualism" *Ethics* 112, 429-470.
- , 2011. "Dispassionate Opprobrium" in the book *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. 2011. Edited by R. Jay Wallace, Rahul Kumar and Samuel Friedman. Oxford: Oxford University Press.
- , 2013. "Rightness and Responsibility" in Coates, Justin & Tognazzini, Neal (eds). 2013 *Blame. Its Nature and Norms*.
- , 2014. "Emotions and Relationships: On a Theme from Strawson" in Shoemaker, D. & Tognazzini N. (eds.), 2014. *Oxford Studies in Agency and Responsibility (Volume 2) 'Freedom and Resentment' at 50*.
- Wallace, R. J., Kumar, R. and Friedman, S. (eds.). 2011. *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford: Oxford University Press.
- Waller, B. 2011. *Against Moral Responsibility*. MIT Press, Cambridge, Massachusetts.
- Watson, Gary 1987. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme" in Schoeman (ed.) 1987.
- , 2004/[2001]. "Reasons and Responsibility". *Ethics* vol. 111 no.1, 374-394. Reprinted in Watson, G. 2004.
- , 2004/[1996]. "Two Faces of Responsibility". Reprinted in Watson, G. 2004 *Agency and Answerability: Selected Essays*.
- , 2004. *Agency and Answerability: Selected Essays*. Clarendon Press: Oxford.
- , 2011. "The Trouble with Psychopaths" in the book *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. 2011. Edited by R. Jay Wallace, Rahul Kumar and Samuel Friedman. Oxford: Oxford University Press.
- , 2013. "Standing in Judgment" in Coates, Justin & Tognazzini, Neal (eds.). 2013 *Blame. Its Nature and Norms*.
- , 2014. "Peter Strawson on Responsibility and Sociality" in Shoemaker, D. & Tognazzini N. (eds.), 2014. *Oxford Studies in Agency and Responsibility (Volume 2) 'Freedom and Resentment' at 50*.
- Williams, Bernard 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- , 1985. *Ethics and the Limits of Philosophy*. London: Fontana.
- , 1993. *Shame and Necessity*. Berkeley: University of California Press.
- , 1997. "Moral Responsibility and Political Freedom". *Cambridge Law Journal* 56(1) March 1997, pp 96-102.
- Williams, Garrath 2008. "Responsibility as a Virtue", *Ethical Theory and Moral Practice*. August 2008, Volume 11, Issue 4, pp. 455-470.
- , 2013. "Sharing Responsibility and Holding Responsible". *Journal of Applied Philosophy*, Vol. 30, No. 4, 2013.
- Winthrop, Delba 1975. "Aristotle and Political Responsibility", *Political Theory*, vol. 3, no. 4 (March 1975).

- Wilson, Fred 2003. "Hume, Treatise of Human Nature – A Genial Skepticism, an Ethical Naturalism", in the book *The Classics of Western Philosophy – A Reader's Guide*, Gracia, Reichberg & Schumacher (toim.), Padstow: Blackwell Publishing.
- Wolf, Susan 1987. "Sanity and the Metaphysics of Responsibility" in Schoeman (ed.) 1987.
- , 1990. *Freedom within Reason*. Oxford: Oxford University Press.
- , 2011. "Blame, Italian Style" in the book *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. 2011. Edited by R. Jay Wallace, Rahul Kumar and Samuel Friedman. Oxford: Oxford University Press.

Internet resources

- Alfano, M. 2013b. "Skepticism about Character" on *philpapers.org* (<http://philpapers.org/browse/skepticism-about-character>). (26.2.2016.)
- Clarke, R & Capes, J. 2013/[2000]. "Incompatibilist (Nondeterministic) Theories of Free Will" *Stanford Encyclopedia of Philosophy*. plato.stanford.edu/entries/incompatibilist-theories (19. 2. 2015).
- Coates, Justin & Tognazzini, Neal (eds). 2014. "Blame". *Stanford Encyclopedia of Philosophy*. (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/blame/>. (20.8.2014).
- Dennett, D. 2011. Review of *Against Moral Responsibility* by Bruce Waller. Published on naturalism.org by MIT press. URL = <http://www.naturalism.org/DCDWallerreview.htm>. (13.4.2015).
- Eshleman, Andrew S. 2014/[2001]. "Moral Responsibility" *Stanford Encyclopedia of Philosophy*. plato.stanford.edu/entries/moral-responsibility (20.8.2014).
- Falcon, Andrea. 2012/[2006]. "Aristotle on Causality" *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/aristotle-causality/>. (25.2.2015).
- Homiak, Marcia 2015/[2003]. "Moral Character" *Stanford Encyclopedia of Philosophy*. plato.stanford.edu/entries/moral-character/ (29.9.2014).
- Honderich, T. (ed.) "The Determinism and Freedom Philosophy Website" www.ucl.ac.uk/~uctytho/dfwIntroIndex.htm (20.8.2014).
- McKenna, M. 2009. "Compatibilism", *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/compatibilism/>. (21.8.2014).
- Parker-Ryan, Sally 2012. "Ordinary Language Philosophy" *Internet Encyclopedia of Philosophy*. URL: <http://www.iep.utm.edu/ord-lang/> Edited by Fieser, J & Dowden, B. (8.6.2015).
- Smiley, M. 2010/[2005]. "Collective Responsibility" *Stanford Encyclopedia of Philosophy*. *Stanford Encyclopedia of Philosophy*. plato.stanford.edu/entries/collective-responsibility/ (21.1.2015). Edward Zalta (ed.).
- Williams, Garrath 2009. "Responsibility", *The Internet Encyclopedia of Philosophy*, <URL: <http://www.iep.utm.edu/responsi/>> Edited by Fieser, J & Dowden, B. (22.5.2013).

–, 2010. "Moral Responsibility", *Oxford Bibliographies Online*, at URL:
<<http://www.oxfordbibliographies.com/view/document/obo-9780195396577/obo-9780195396577-0079.xml?rskey=XtPLXC&result=1&q=Moral%20Responsibility#firstMatch>>
(22.5.2013).

Conclusion